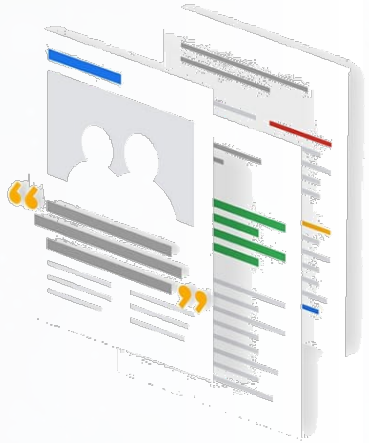
DTSV



Data Storytelling & Visualization

13 October 2020

Powered by Google Data Studio

Organized by Universiti Teknologi MARA | Pejabat Pembangunan Infrastruktur dan Infostruktur

INFOSTRUKTUR WEBINAR 2020

Please Scan **QR Code** for **Participant Profile Survey**

Agenda

- About This Webinar
- Data Storytelling & Visualization
- Google Data Studio + Demo
- Q&A
- DSTV Quiz
- Get Google Data Studio Certificate
- Feedback Form & Attendance

Data Storytelling

Welcome to your story.

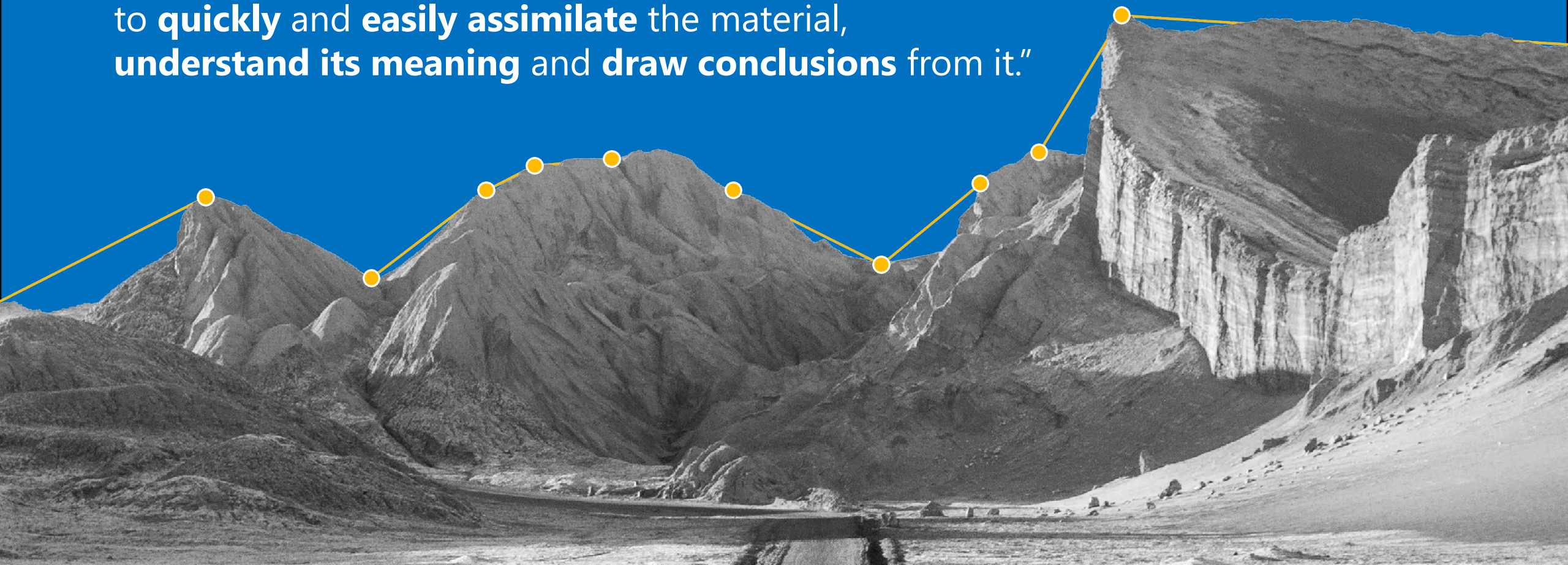
Every **story is unique**. It's an opportunity to **inspire**, **advise** or **enlighten**.
Today you will discover the story hidden within your data, so you can become the author that ignites **imaginations** and **turns information into knowledge**.

How to create **compelling data visualizations?**
In this knowledge sharing session you will discover:

- **how to approach your data**,
- **how to make the most of elements within data storytelling**, and
- **decide which visualizations tell your audience the best story.**

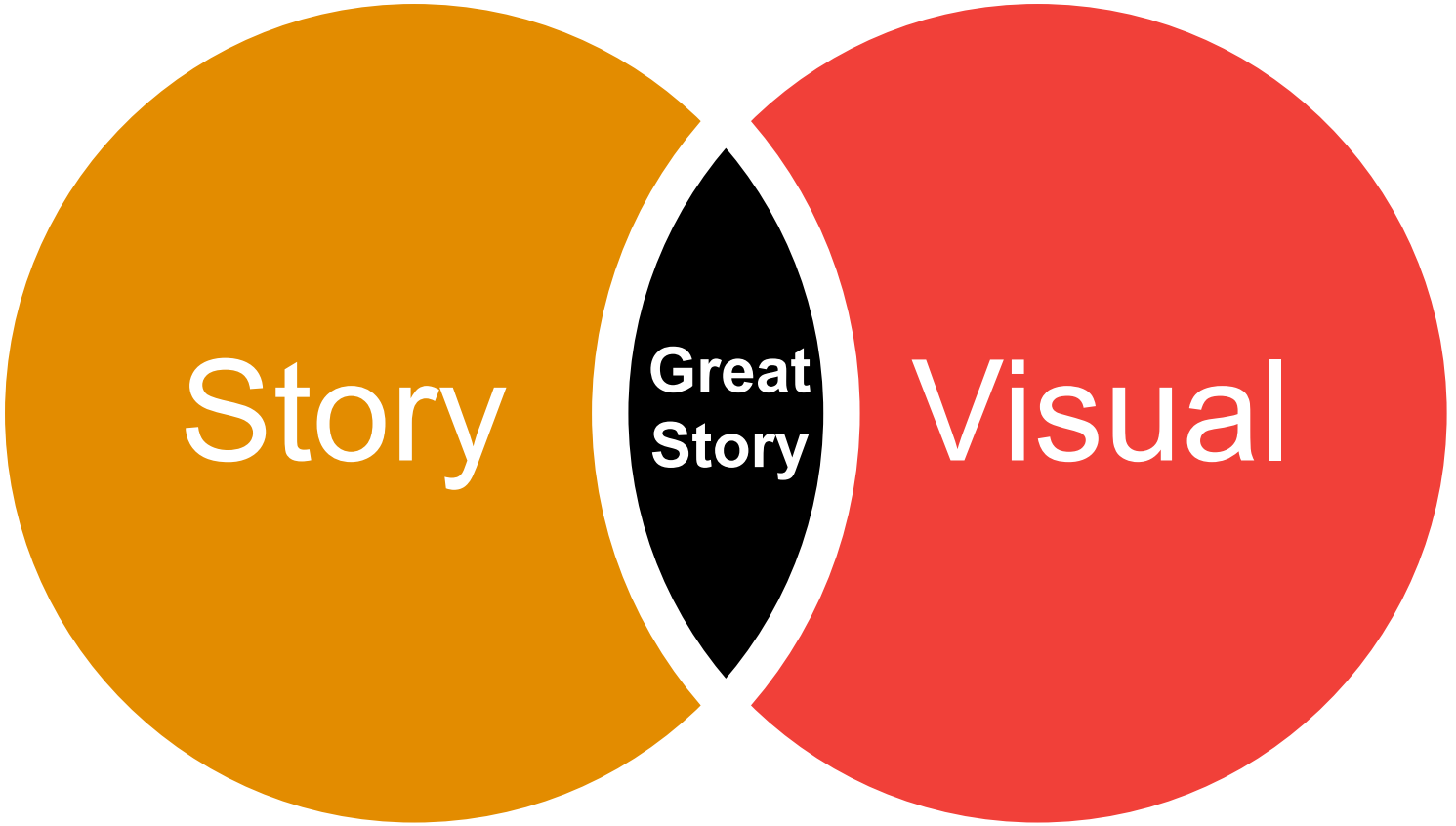
Data Storytelling:
Definition

"A **method** of **delivering messages** derived from complex **data** analysis in a way that allows the audience to **quickly** and **easily assimilate** the material, **understand its meaning** and **draw conclusions** from it."



Two parts to a great Data Story

One half alone will not get you anywhere!



Data Storytelling:

Data storytelling involves:

- **understanding**, **processing**, and **extracting data**
- then **creating visuals** to display the data using an **understandable picture;** and
- followed by **communicating** the **data story with a narrative**.

These are the **three key elements** to communicating information well.

Source: Forbes

Data Storytelling:
The Process

1st Step: **Data** - example Employee, Student, Research, Financials and other subject areas feeding information into Analytics.

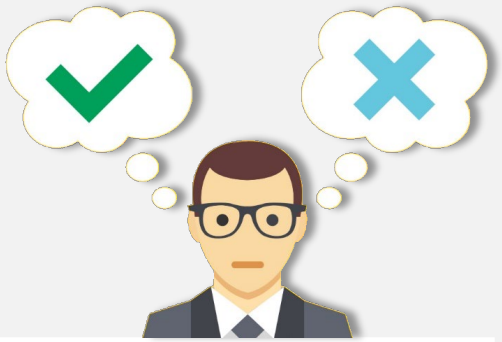
2nd Step: In Analytics, we can create **visual** graphs to summarize and distill the data.

3rd step: **Narrative** or what is called data storytelling – it is coming up with the story to *highlight* the *key points*, *guide* the *questions* that need to be asked, and help *inform* the *actions* that are needed.

Source: Forbes

Data Storytelling:
The Goal

The goal of data storytelling is **to support the decision making process** and to **facilitate changes** so the organization can achieve it's strategic goals.



STRATEGIC GOALS

Source: Forbes


Visual Storytelling?
Data professionals without it have lesser impact.
25%
of Data Scientists report that the work they do is **not used by decision-makers.**
22%
of Data Scientists **struggle to explain** Data Science to others.

Bad Storytelling

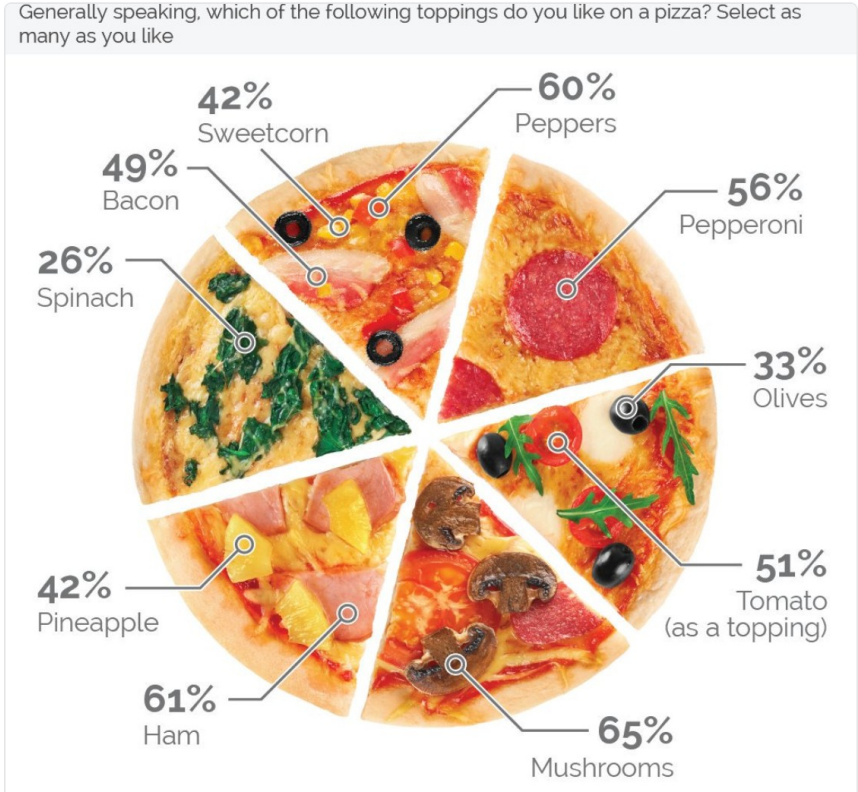
Notice that most of today's examples will have **good** visuals but a **bad/non-existent** story

Beautiful Visual – but what exactly is the story here?

Hint: Notice the slices!

YouGov
@YouGov
Follow

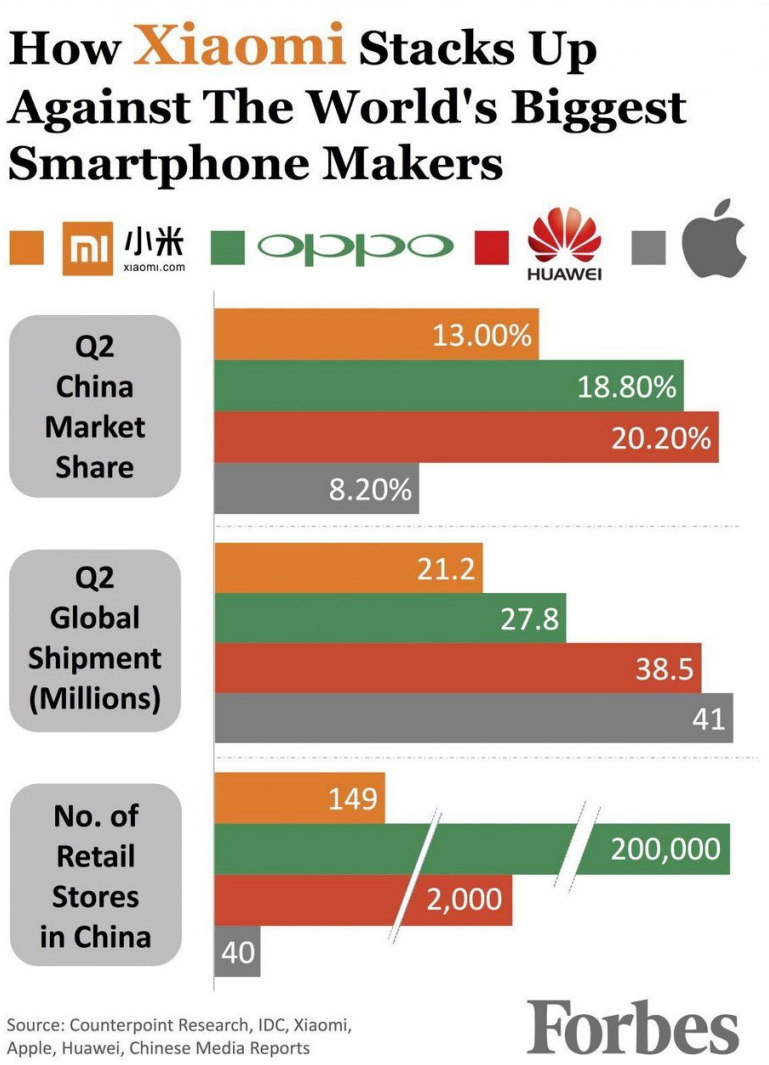
Forget pepperoni - mushroom is Britain's most liked pizza topping (65%), followed by onion (62%) and then ham (61%)
yougov.co.uk/news/2017/03/0 …



Source: https://twitter.com/yougov/status/838720989991223297?lang=en

Just because we CAN fit everything on one page..

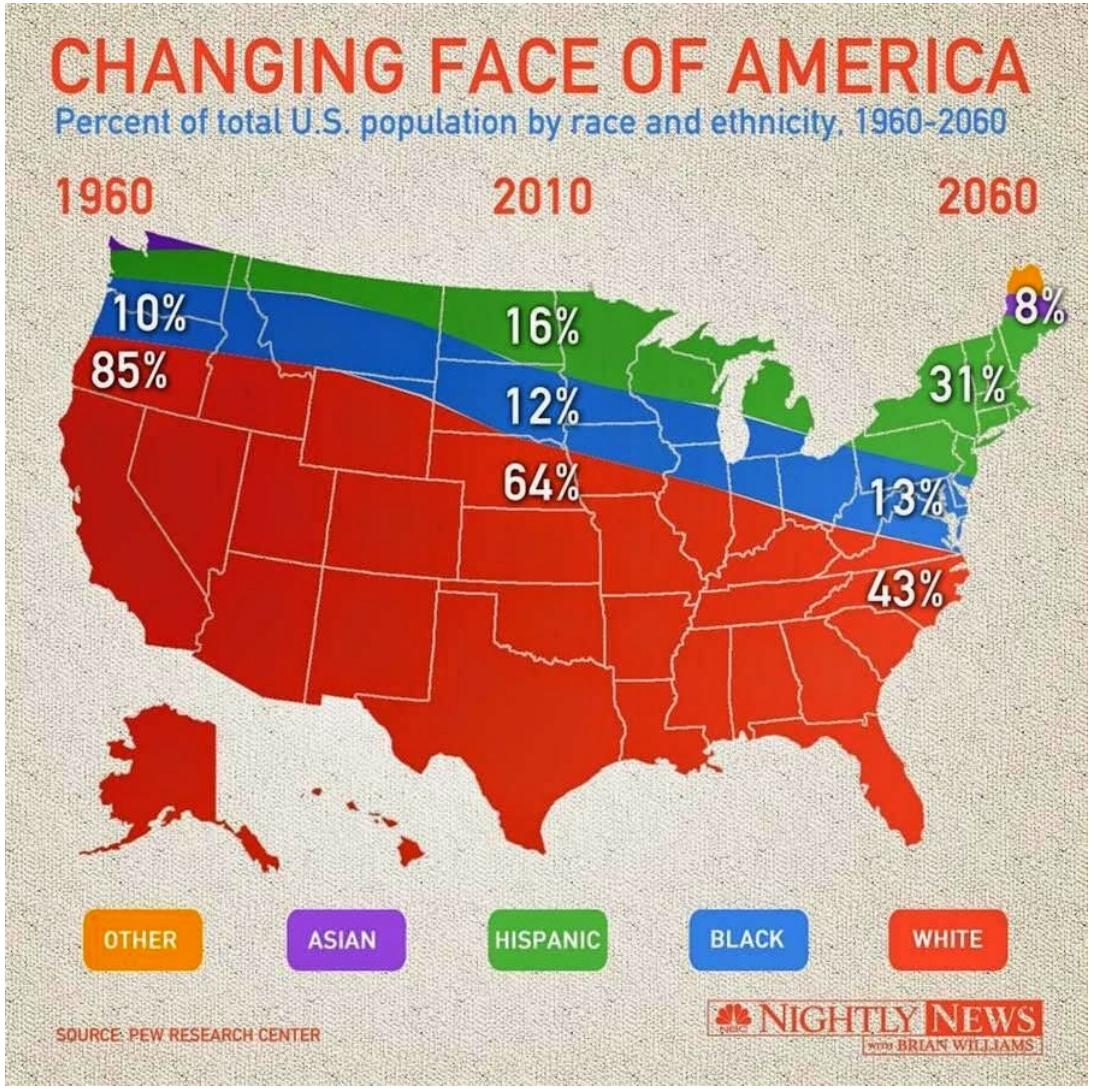
Doesn’t mean you SHOULD



Source: https://www.forbes.com/sites/ywang/2017/08/08/xiaomi-is-once-again-worlds-biggest-smartphone-brand-but-could-it-go-further/#5e7637e51691

Distorted Geographic Map

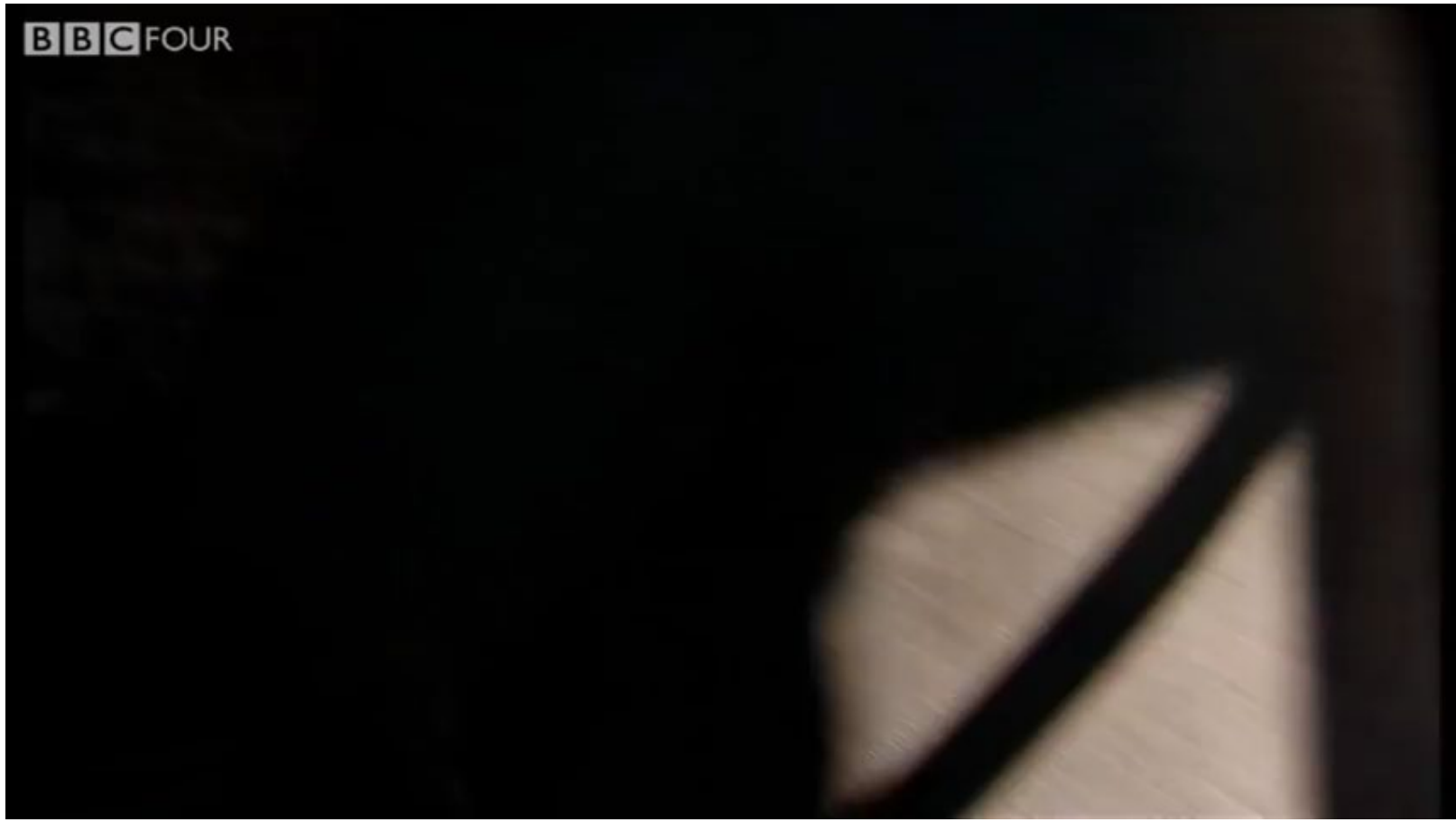
How would you interpret this map of America?



Source: http://livingqlikview.com/the-9-worst-data-visualizations-ever-created/

A Good Data Story combines BOTH Story & Visual

Show the numbers, but with purpose!



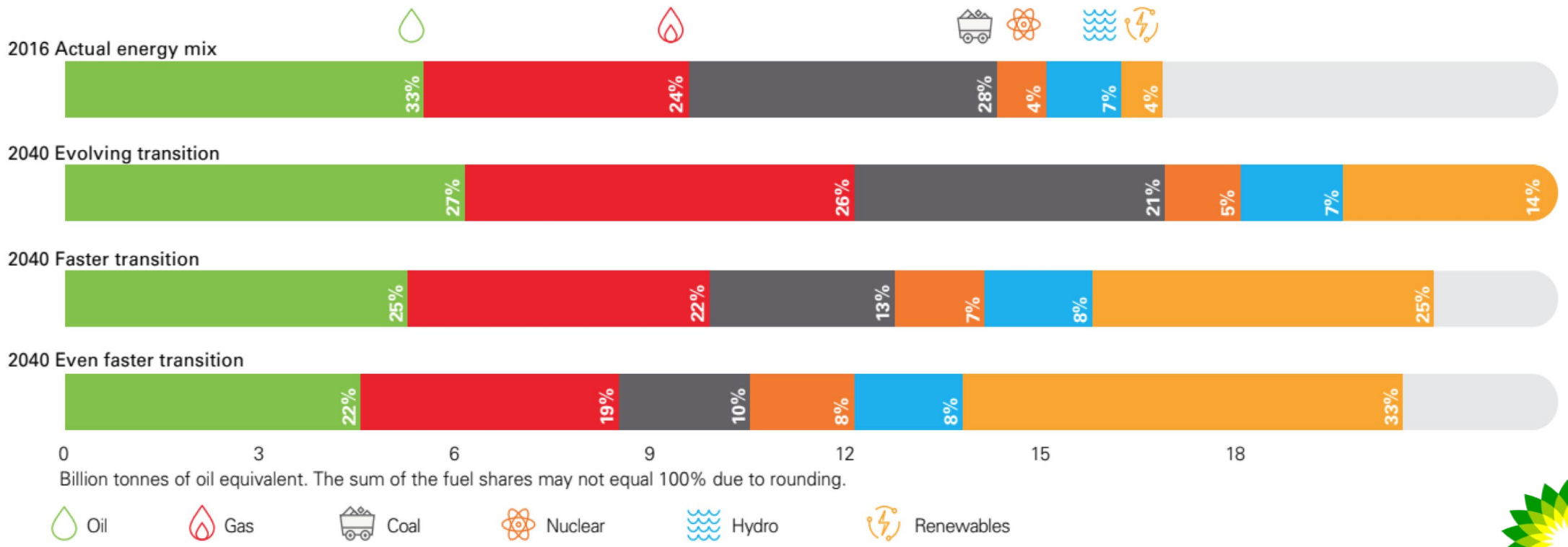
Oil & Gas Specific Examples?

What data stories do other oil & gas companies have to share?

What is this visual trying to tell us?

BP 2017 Annual Report

Energy consumption – 2040 projections

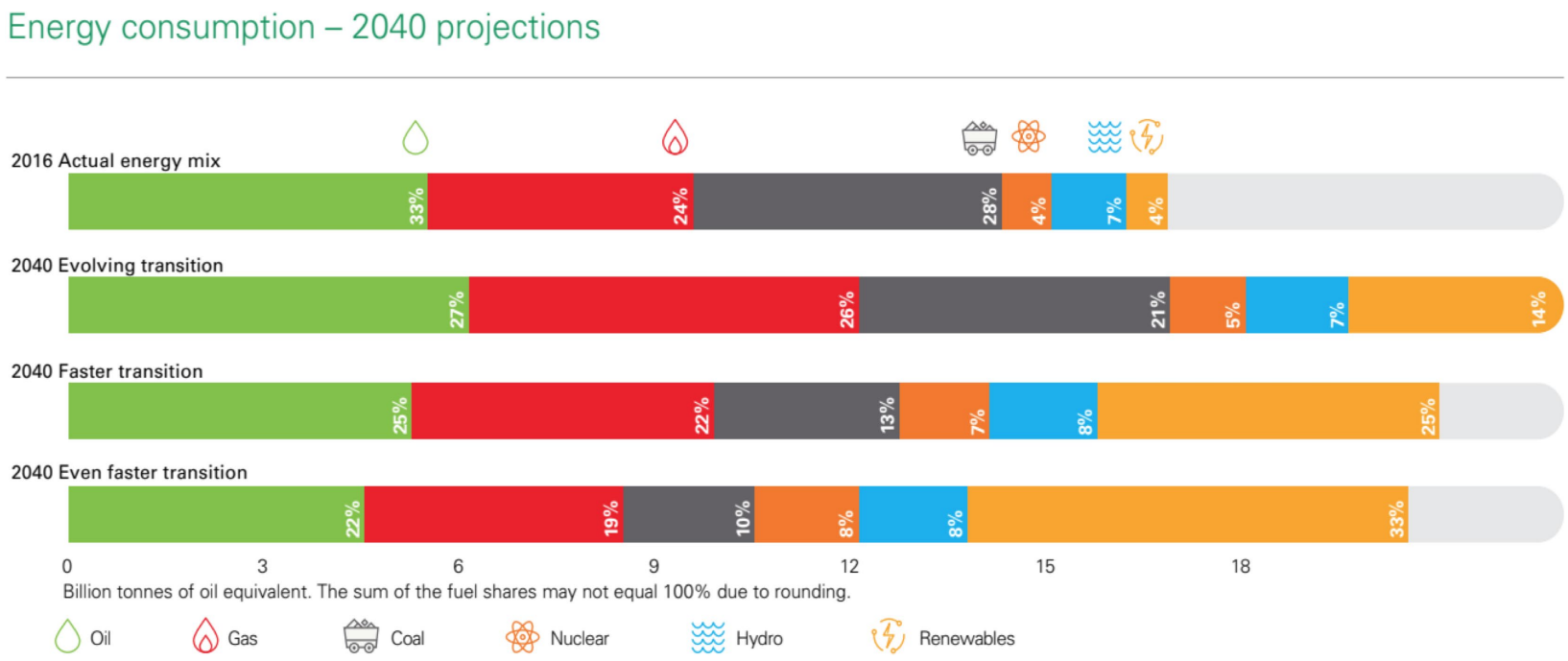


bp

Beautiful Visual without a cohesive story

Too much information, too small a space

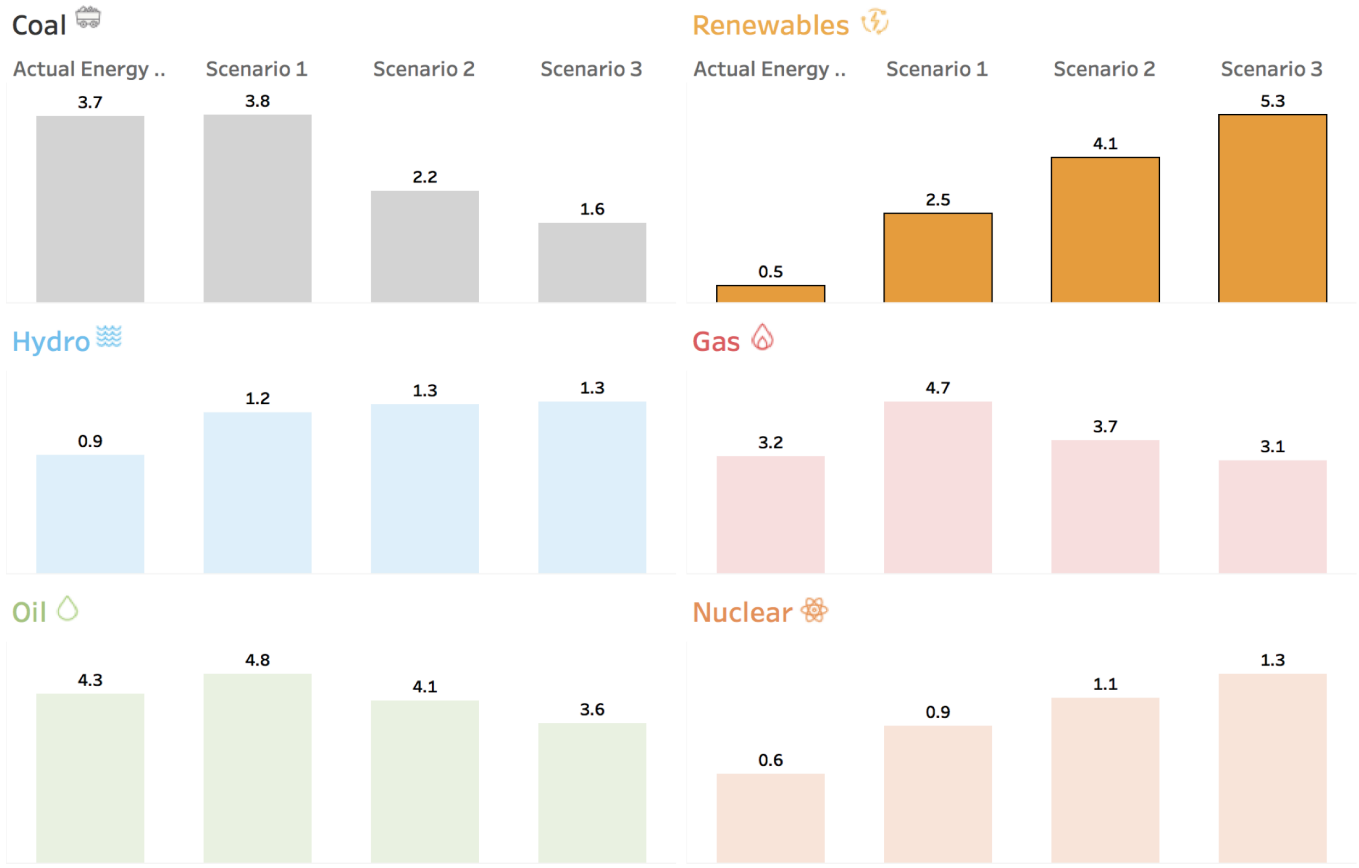
1. Cognitive Overload
2. Uninformative Title
3. Proportions not comparable
4. Axis not clear



Suggested re-design – Refine the message

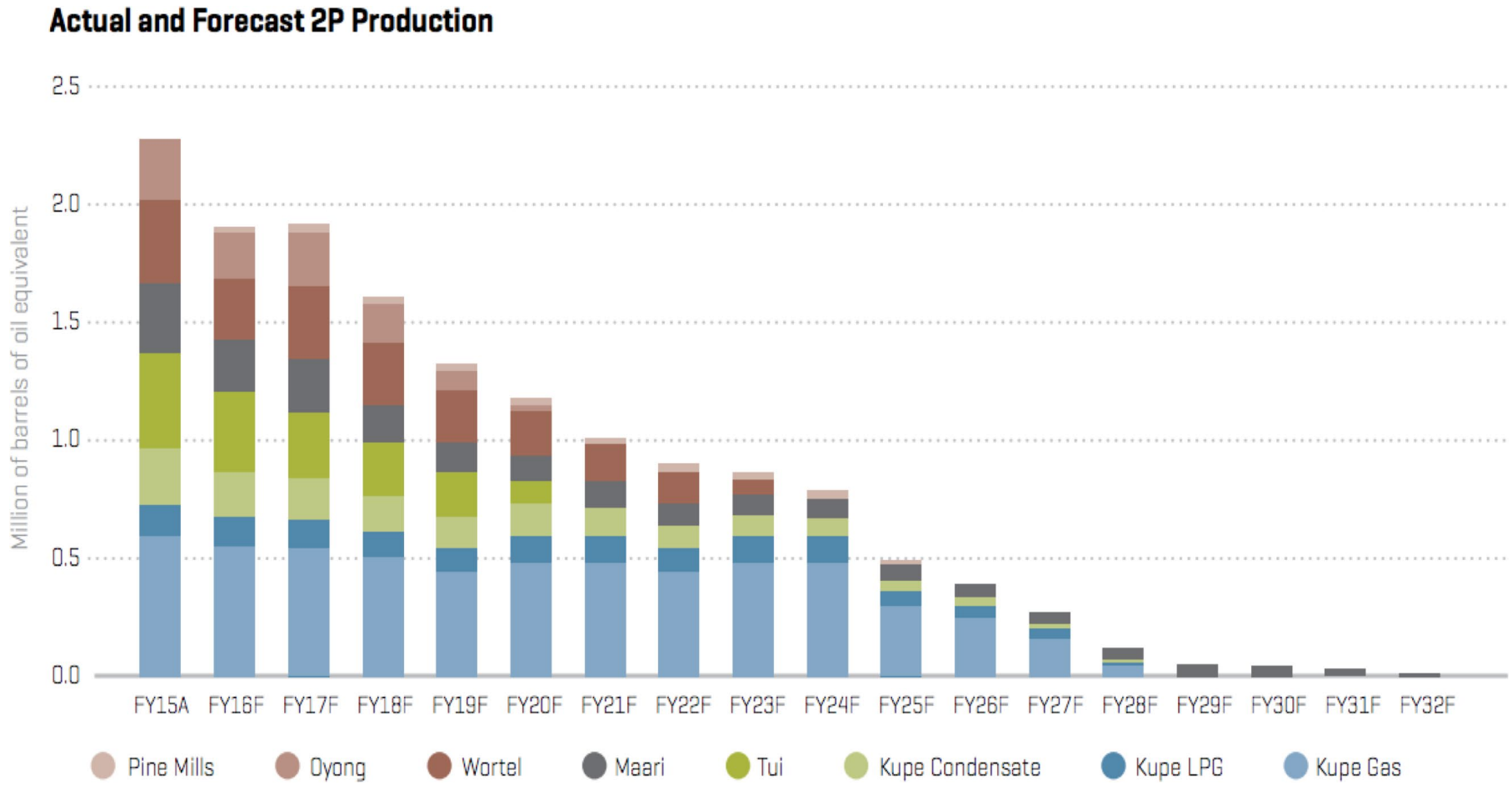
Make ‘Renewables’ the focus of the visual & build a visual hierarchy.

We anticipate **Renewables** to be a growing energy source in all three scenarios of the future. This is in response to both a growing global appetite for energy as well as a reduced reliance on traditional energy sources such as **Coal**, **Gas** & **Oil**.



What is this visual trying to tell us?

New Zealand Oil & Gas 2015 Annual Report

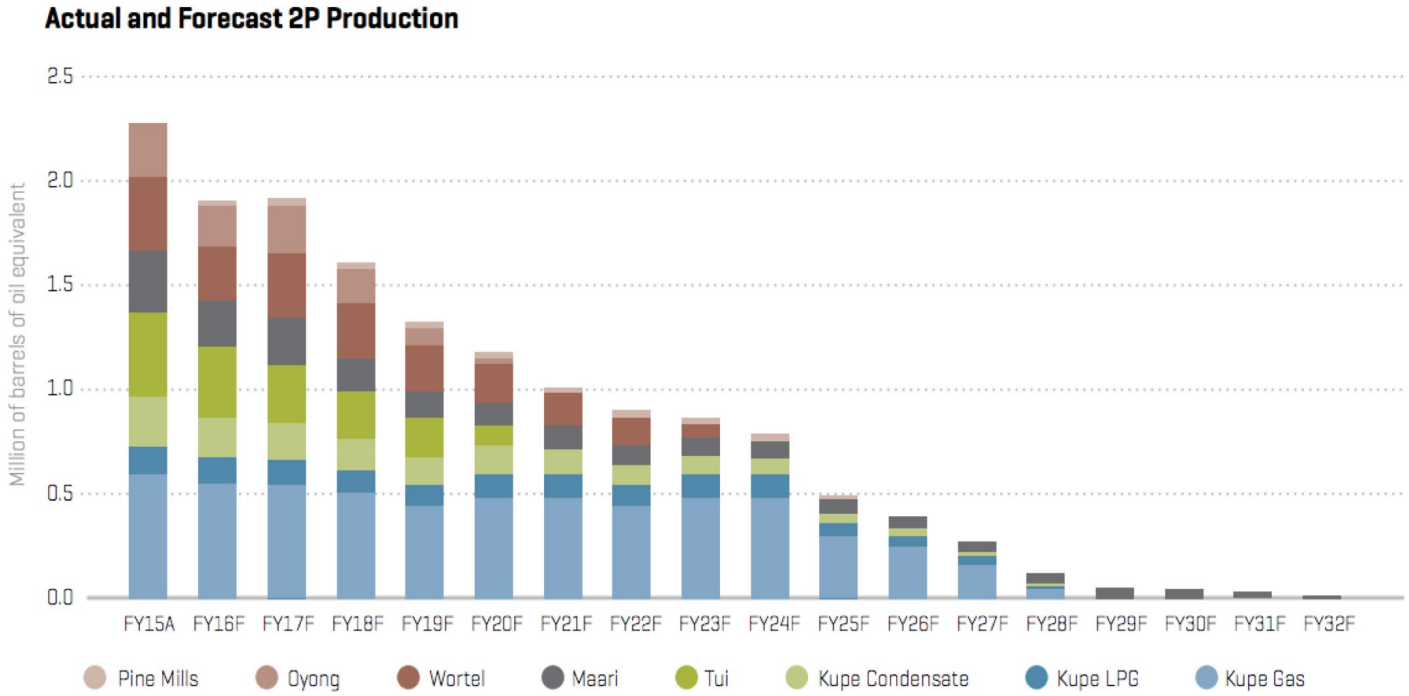


NEW ZEALAND OIL & GAS

What is this visual trying to tell us?

New Zealand Oil & Gas 2015 Annual Report

1. Segments not comparable
2. Uninformative Title
3. Confusing color scheme
4. Wasted space on axes

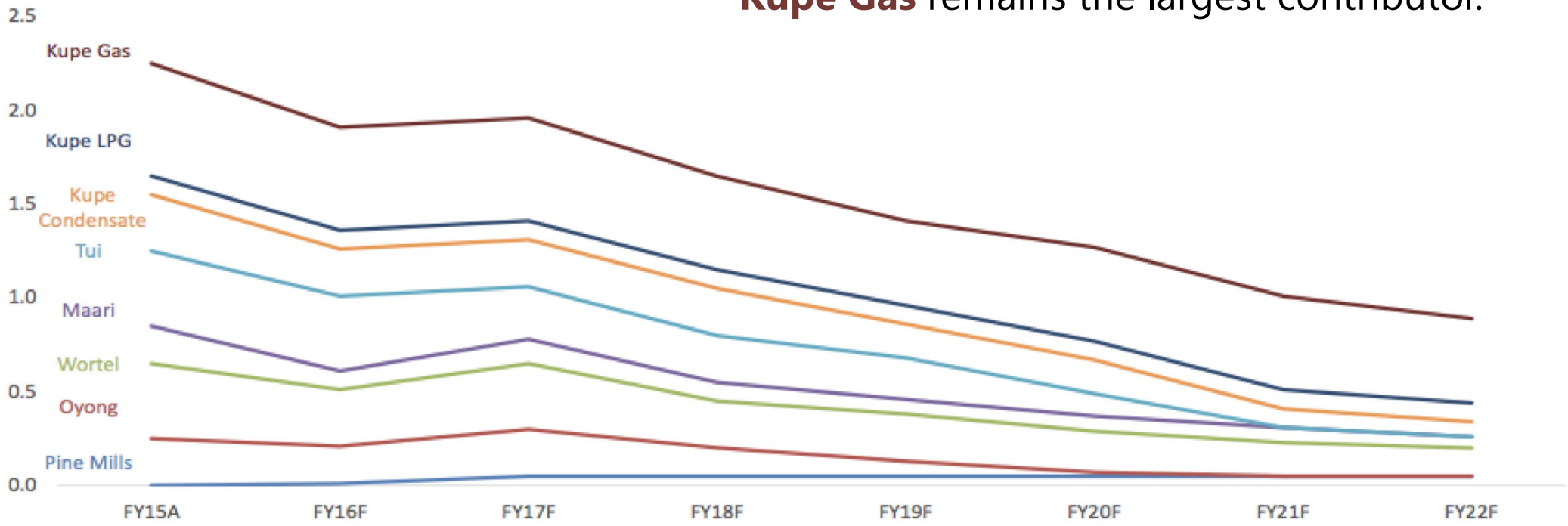


NEW ZEALAND OIL & GAS

Suggested re-design – Choose the correct chart time

Line charts are good for change over time – Note we should remove excess color!

New Zealand Oil & Gas anticipate falling production across all their old fields. Despite declining production, **Kupe Gas** remains the largest contributor.



NEW ZEALAND OIL & GAS

How did thishappen?

Let’s put ourselves in the shoes of an Analyst.

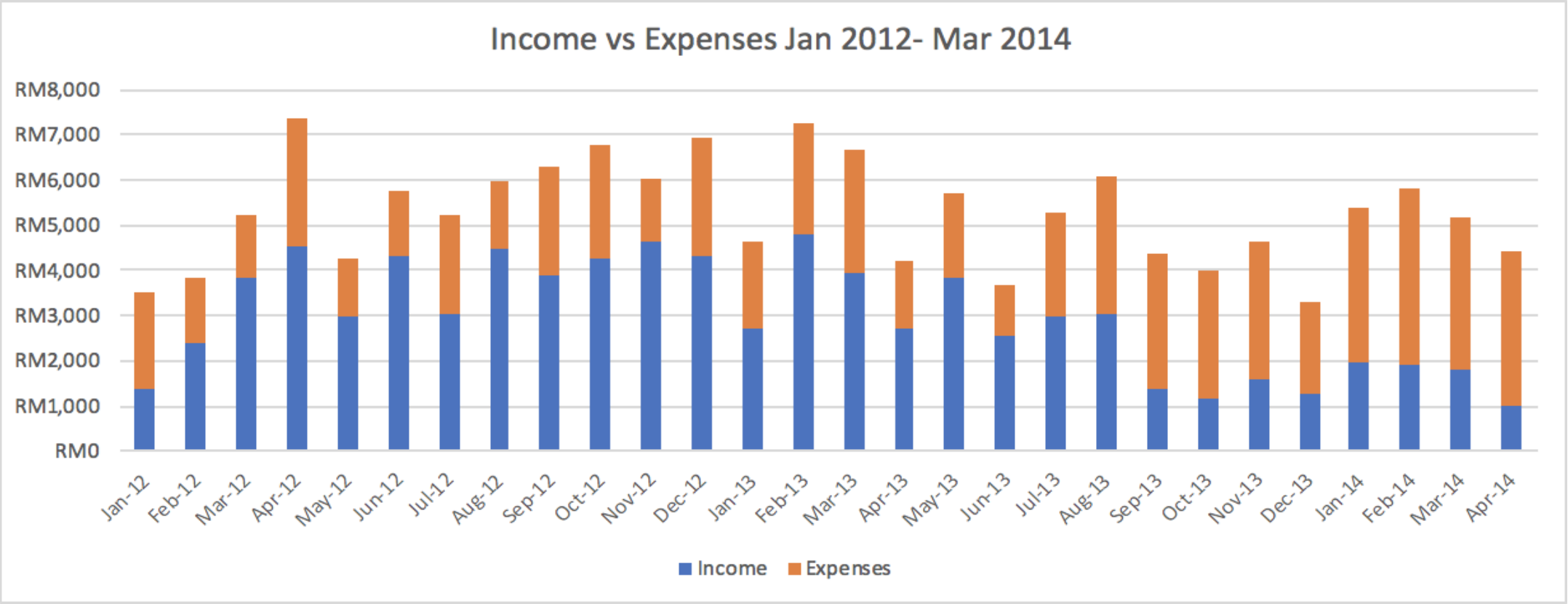
How many people work with data that looks like this?

Tabular data is very common – almost all of us start here.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Profit Ratio	Category	City	Country	Customer Na	Discount	Number of R	Order Date	Order ID	Postal Code	Manufacture	Product Nan	Profit	Quantity	Region	Sales	Segment	Ship Date	Ship Mode	State	Sub-Categor
2	16%	Furniture	Henderson	United State	Claire Gute	0%	1	08/11/2017	CA-2017-152	42420	Bush	Bush Somers	$42	2	South	$262	Consumer	11/11/2017	Second Class	Kentucky	Bookcases
3	30%	Furniture	Henderson	United State	Claire Gute	0%	1	08/11/2017	CA-2017-152	42420	Hon	Hon Deluxe I	$220	3	South	$732	Consumer	11/11/2017	Second Class	Kentucky	Chairs
4	47%	Office Suppli	Los Angeles	United State	Darrin Van H	0%	1	12/06/2017	CA-2017-138	90036	Universal	Self-Adhesive	$7	2	West	$15	Corporate	16/06/2017	Second Class	California	Labels
5	-40%	Furniture	Fort Lauderc	United State	Sean O'Donn	45%	1	11/10/2016	US-2016-108	33311	Bretford	Bretford CR4	-$383	5	South	$958	Consumer	18/10/2016	Standard Cla	Florida	Tables
6	11%	Office Suppli	Fort Lauderc	United State	Sean O'Donn	20%	1	11/10/2016	US-2016-108	33311	Eldon	Eldon Fold 'N	$3	2	South	$22	Consumer	18/10/2016	Standard Cla	Florida	Storage
7	29%	Furniture	Los Angeles	United State	Brosina Hoff	0%	1	09/06/2015	CA-2015-115	90032	Eldon	Eldon Expres	$14	7	West	$49	Consumer	14/06/2015	Standard Cla	California	Furnishings
8	27%	Office Suppli	Los Angeles	United State	Brosina Hoff	0%	1	09/06/2015	CA-2015-115	90032	Newell	Newell 322	$2	4	West	$7	Consumer	14/06/2015	Standard Cla	California	Art
9	10%	Technology	Los Angeles	United State	Brosina Hoff	20%	1	09/06/2015	CA-2015-115	90032	Mitel	Mitel 5320 I	$91	6	West	$907	Consumer	14/06/2015	Standard Cla	California	Phones
10	31%	Office Suppli	Los Angeles	United State	Brosina Hoff	20%	1	09/06/2015	CA-2015-115	90032	DXL	DXL Angle-Vi	$6	3	West	$19	Consumer	14/06/2015	Standard Cla	California	Binders
11	30%	Office Suppli	Los Angeles	United State	Brosina Hoff	0%	1	09/06/2015	CA-2015-115	90032	Belkin	Belkin F5C2(	$34	5	West	$115	Consumer	14/06/2015	Standard Cla	California	Appliances
12	5%	Furniture	Los Angeles	United State	Brosina Hoff	20%	1	09/06/2015	CA-2015-115	90032	Chromcraft	Chromcraft I	$85	9	West	$1,706	Consumer	14/06/2015	Standard Cla	California	Tables
13	8%	Technology	Los Angeles	United State	Brosina Hoff	20%	1	09/06/2015	CA-2015-115	90032	Other	Konftel 250 (	$68	4	West	$911	Consumer	14/06/2015	Standard Cla	California	Phones
14	35%	Office Suppli	Concord	United State	Andrew Aller	20%	1	15/04/2018	CA-2018-114	28027	Xerox	Xerox 1967	$5	3	South	$16	Consumer	20/04/2018	Standard Cla	North Caroli	Paper
15	33%	Office Suppli	Seattle	United State	Irene Maddo	20%	1	05/12/2017	CA-2017-161	98103	Fellowes	Fellowes PB2	$133	3	West	$408	Consumer	10/12/2017	Standard Cla	Washington	Binders
16	-180%	Office Suppli	Fort Worth	United State	Harold Pawl	80%	1	22/11/2016	US-2016-118	76106	Holmes	Holmes Repl	-$124	5	Central	$69	Home Office	26/11/2016	Standard Cla	Texas	Appliances
17	-150%	Office Suppli	Fort Worth	United State	Harold Pawl	80%	1	22/11/2016	US-2016-118	76106	Storex	Storex Dura1	-$4	3	Central	$3	Home Office	26/11/2016	Standard Cla	Texas	Binders
18	2%	Office Suppli	Madison	United State	Pete Kriz	0%	1	11/11/2015	CA-2015-105	53711	Other	Stur-D-Stor S	$13	6	Central	$666	Consumer	18/11/2015	Standard Cla	Wisconsin	Storage
19	18%	Office Suppli	West Jordan	United State	Alejandro Gr	0%	1	13/05/2015	CA-2015-167	84084	Fellowes	Fellowes Sup	$10	2	West	$56	Consumer	15/05/2015	Second Class	Utah	Storage
20	29%	Office Suppli	San Francisc	United State	Zuschuss Do	0%	1	27/08/2015	CA-2015-143	94109	Newell	Newell 341	$2	2	West	$9	Consumer	01/09/2015	Second Class	California	Art
21	7%	Technology	San Francisc	United State	Zuschuss Do	20%	1	27/08/2015	CA-2015-143	94109	Cisco	Cisco SPA 50	$16	3	West	$213	Consumer	01/09/2015	Second Class	California	Phones
22	33%	Office Suppli	San Francisc	United State	Zuschuss Do	20%	1	27/08/2015	CA-2015-143	94109	Wilson Jone	Wilson Jone	$7	4	West	$23	Consumer	01/09/2015	Second Class	California	Binders
23	26%	Office Suppli	Fremont	United State	Ken Black	0%	1	09/12/2017	CA-2017-137	68025	Newell	Newell 318	$5	7	Central	$19	Corporate	13/12/2017	Standard Cla	Nebraska	Art
24	26%	Office Suppli	Fremont	United State	Ken Black	0%	1	09/12/2017	CA-2017-137	68025	Acco	Acco Six-Out	$16	7	Central	$60	Corporate	13/12/2017	Standard Cla	Nebraska	Appliances
25	-1%	Furniture	Philadelphia	United State	Sandra Flana	30%	1	16/07/2018	US-2018-156	19140	Global	Global Delu>	-$1	2	East	$71	Consumer	18/07/2018	Second Class	Pennsylvania	Chairs
26	23%	Furniture	Orem	United State	Emily Burns	0%	1	25/09/2016	CA-2016-106	84057	Bretford	Bretford CR4	$240	3	West	$1,045	Consumer	30/09/2016	Standard Cla	Utah	Tables
27	36%	Office Suppli	Los Angeles	United State	Eric Hoffmar	20%	1	16/01/2017	CA-2017-121	90049	Wilson Jone	Wilson Jone	$4	2	West	$12	Consumer	20/01/2017	Second Class	California	Binders
28	13%	Technology	Los Angeles	United State	Eric Hoffmar	0%	1	16/01/2017	CA-2017-121	90049	Other	Imation 8GB	$12	3	West	$91	Consumer	20/01/2017	Second Class	California	Accessories
29	-54%	Furniture	Philadelphia	United State	Tracy Blumst	50%	1	17/09/2016	US-2016-150	19140	Riverside	Riverside Pal	-$1,665	7	East	$3,083	Consumer	21/09/2016	Standard Cla	Pennsylvania	Bookcases
30	-73%	Office Suppli	Philadelphia	United State	Tracy Blumst	70%	1	17/09/2016	US-2016-150	19140	Avery	Avery Recycl	-$7	2	East	$10	Consumer	21/09/2016	Standard Cla	Pennsylvania	Binders
31	13%	Furniture	Philadelphia	United State	Tracy Blumst	20%	1	17/09/2016	US-2016-150	19140	Howard Mill	Howard Mill	$16	3	East	$124	Consumer	21/09/2016	Standard Cla	Pennsylvania	Furnishings
32	34%	Office Suppli	Philadelphia	United State	Tracy Blumst	20%	1	17/09/2016	US-2016-150	19140	Poly	Poly String T	$1	2	East	$3	Consumer	21/09/2016	Standard Cla	Pennsylvania	Envelopes
33	11%	Office Suppli	Philadelphia	United State	Tracy Blumst	20%	1	17/09/2016	US-2016-150	19140	Boston	BOSTON Mo(	$10	6	East	$86	Consumer	21/09/2016	Standard Cla	Pennsylvania	Art
34	-83%	Office Suppli	Philadelphia	United State	Tracy Blumst	70%	1	17/09/2016	US-2016-150	19140	Acco	Acco Pressbc	-$6	6	East	$7	Consumer	21/09/2016	Standard Cla	Pennsylvania	Binders
35	23%	Office Suppli	Philadelphia	United State	Tracy Blumst	20%	1	17/09/2016	US-2016-150	19140	Other	Lumber Cray	$4	2	East	$16	Consumer	21/09/2016	Standard Cla	Pennsylvania	Art
36	34%	Office Suppli	Houston	United State	Matt Abelma	20%	1	19/10/2018	CA-2018-107	77095	Easy-staple p	Easy-staple p	$10	3	Central	$29	Home Office	23/10/2018	Second Class	Texas	Paper
37	11%	Technology	Richardson	United State	Gene Hale	20%	1	08/12/2017	CA-2017-117	75080	GE	GE 30524EE(	$123	7	Central	$1,098	Corporate	10/12/2017	First Class	Texas	Phones
38	-78%	Furniture	Richardson	United State	Gene Hale	60%	1	08/12/2017	CA-2017-117	75080	Electrix	Electrix Arch	-$148	5	Central	$191	Corporate	10/12/2017	First Class	Texas	Furnishings

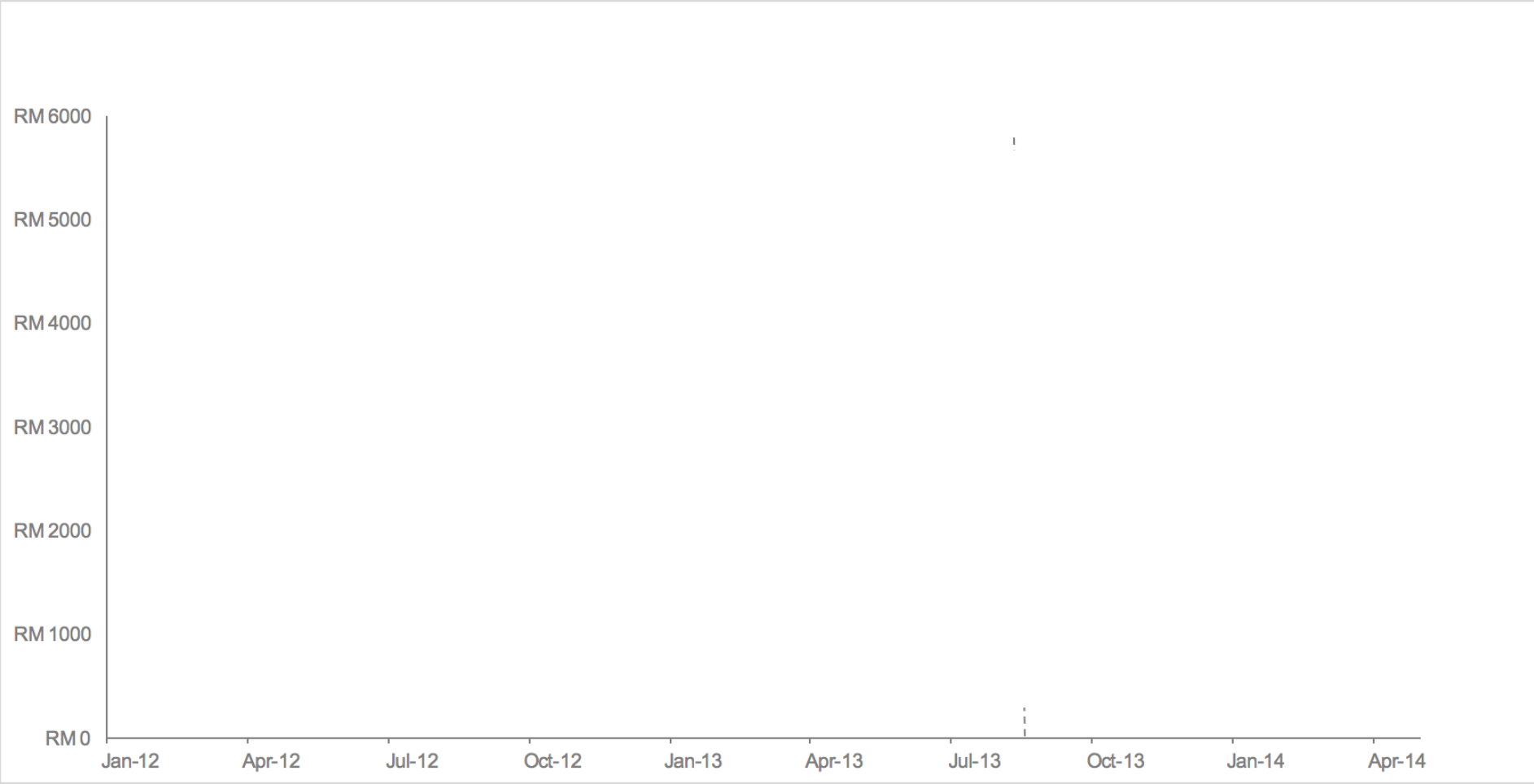
What exactly is this analyst trying to tell us?

We MUST think of the story before plotting visuals



My Story: Expenses > Income because of Strategy Change

Help your decision-makers understand the value of data



The Who, What, How method

Build a profile of your audience,
think about **what to tell them**,
consider **how you can tell them** this story.

Example - Who is our audience?

Vice-President of Strategy at our Toy Company

1. What are they like?

Older / Previously from technical background but has been more client facing recently / Detail-oriented / Time-poor

2. Why are they here?

She is looking for a solution to our problem of falling income & rising expenses

3. What keeps them up at night?

She is new to the role and needs a win / Company Shareholders are critical of her ability / Operational losses unsustainable in long-run

4. How can you solve their problem?

Suggest a viable business strategy to curb losses / Back strategy up with appropriate data

5. How can you best reach them?

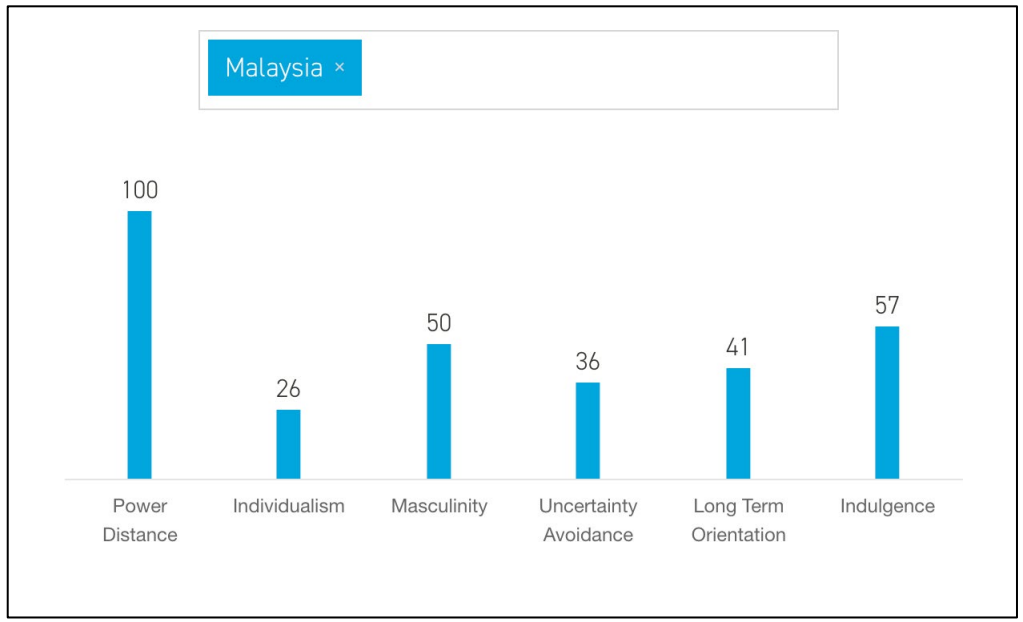
Dislikes lengthy meetings / Tends to rush out last-minute / PowerPoint with Executive Summary & appendix will be good

6. How might they resist?

The proposal must fit within budgetary constraints / needs to be done with readily-available resources

Understand your audience from a cultural perspective

Hofstede's Cultural Dimensions



Source: https://www.hofstede-insights.com/country-comparison/malaysia/

Nationality can have a **big effect on the story/visuals we create**

Malaysia is a **high Power Distance** – we accept that decisions are made at the top

Only need to **persuade** the **senior management**
- no need to persuade entire audience

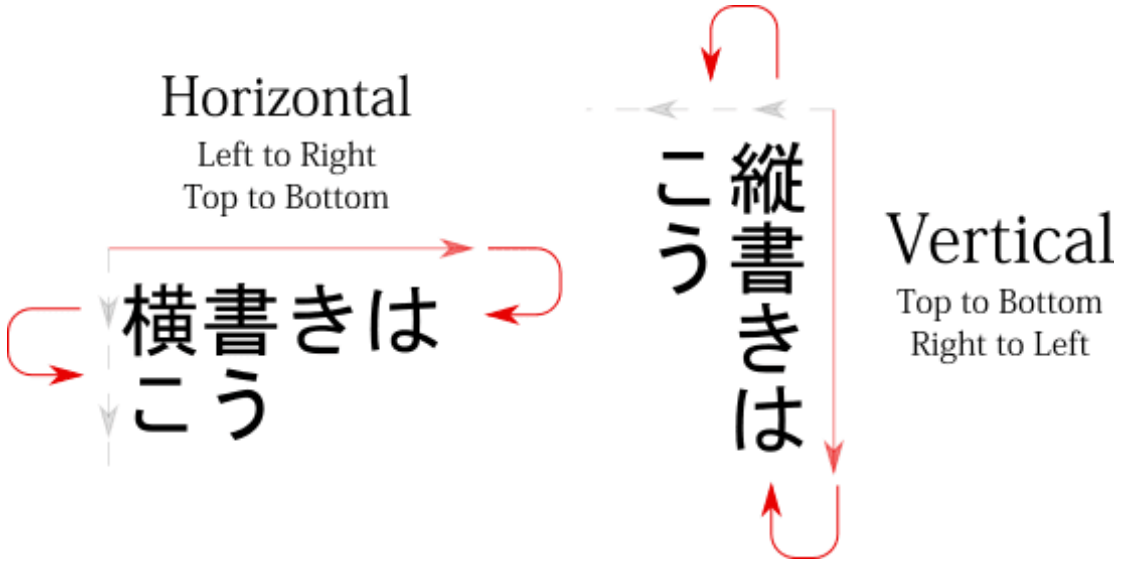
We all have different associations for things

Color associations differ depending on the culture

Source: https://www.six-degrees.com/an-international-guide-on-the-use-of-color-in-marketing-advertising/

The way we read is also influenced by culture

How will their eyes naturally process a page?



٩٢. سورة الليل بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ وَالَّيْلِ إِذَا
يَغْشَى وَالنَّهَارِ إِذَا تَجَلَّى وَمَا خَلَقَ الذَّكَرَ وَالْأُنْثَى إِنَّ
سَعْيَكُمْ لَشَتَّى فَأَمَّا مَنْ أَعْطَى وَاتَّقَى وَصَدَّقَ
بِالْحُسْنَى فَسَنُيَسرُهُ لِلْيُسْرَى وَأَمَّا مَنْ بَخِلَ
وَاسْتَغْنَى وَكَذَّبَ بِالْحُسْنَى فَسَنُيَسِّرُهُ لِلْعُسْرَى وَمَا
يُغْنِي عَنْهُ مَالُهُ إِذَا تَرَدَّى إِنَّ عَلَيْنَا لَلْهدَى وَإِنَّ لَنَا
لَلْاءخِرَةَ وَالْأُولَى فَأَنْذَرْتُكُمْ نَارًا تَلَظَّى لَا يَصْلَهَا إِلَّا
الْأَشْقَى الَّذِي كَذَّبَ وَتَوَلَّى وَسَيُجَنَّبُهَا الْأَتْقَى الَّذِي يُؤْتِي
مَالَهُ يَتَزَكَّى وَمَا لِأَحَدٍ عِنْدَهُ مِنْ نِعْمَةٍ تُجْزَى إِلَّا ابْتِغَاءَ
وَجْهِ رَبِّهِ الْأَعْلَى وَلَسَوْفَ يَرْضَى

Now I know who the audience is. What do I want to tell them?

Craft a refined message.

Analysts like to show the audience ALL their data

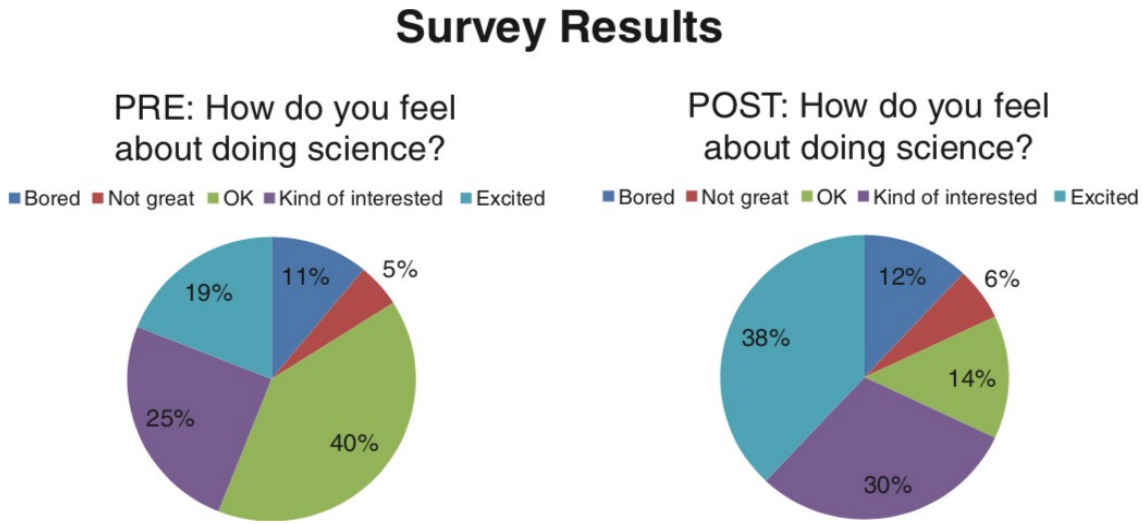
But as the audience we have to ask – so what?

Student Interest Science Before & After attending a Summer Science Camp

Table

Interest	Before	After
Excited	19	38
Kind of interested	25	30
OK	40	14
Not great	5	6
Bored	11	12

Source: Storytelling with Data pg. 5



…. So what?

So . what?

Well… what do they want to get out of your data story?

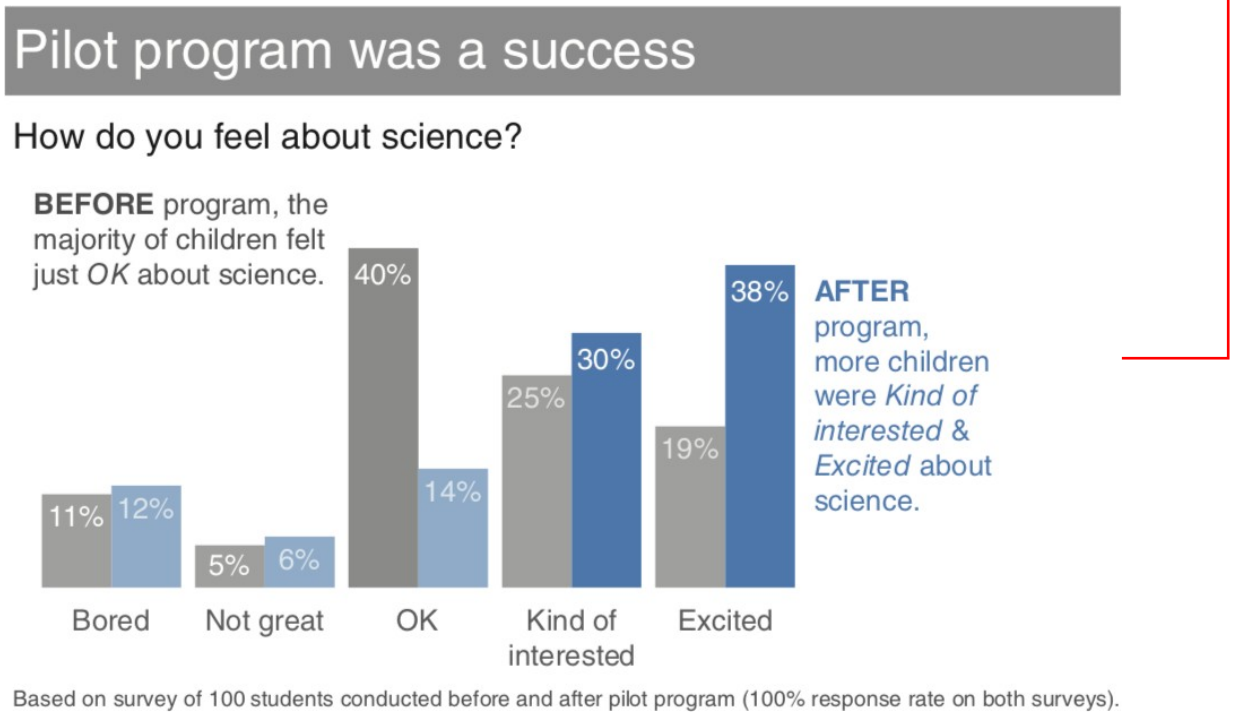
Student Interest Science Before & After attending a Summer Science Camp

Table

Interest	Before	After
Excited	19	38
Kind of interested	25	30
OK	40	14
Not great	5	6
Bored	11	12

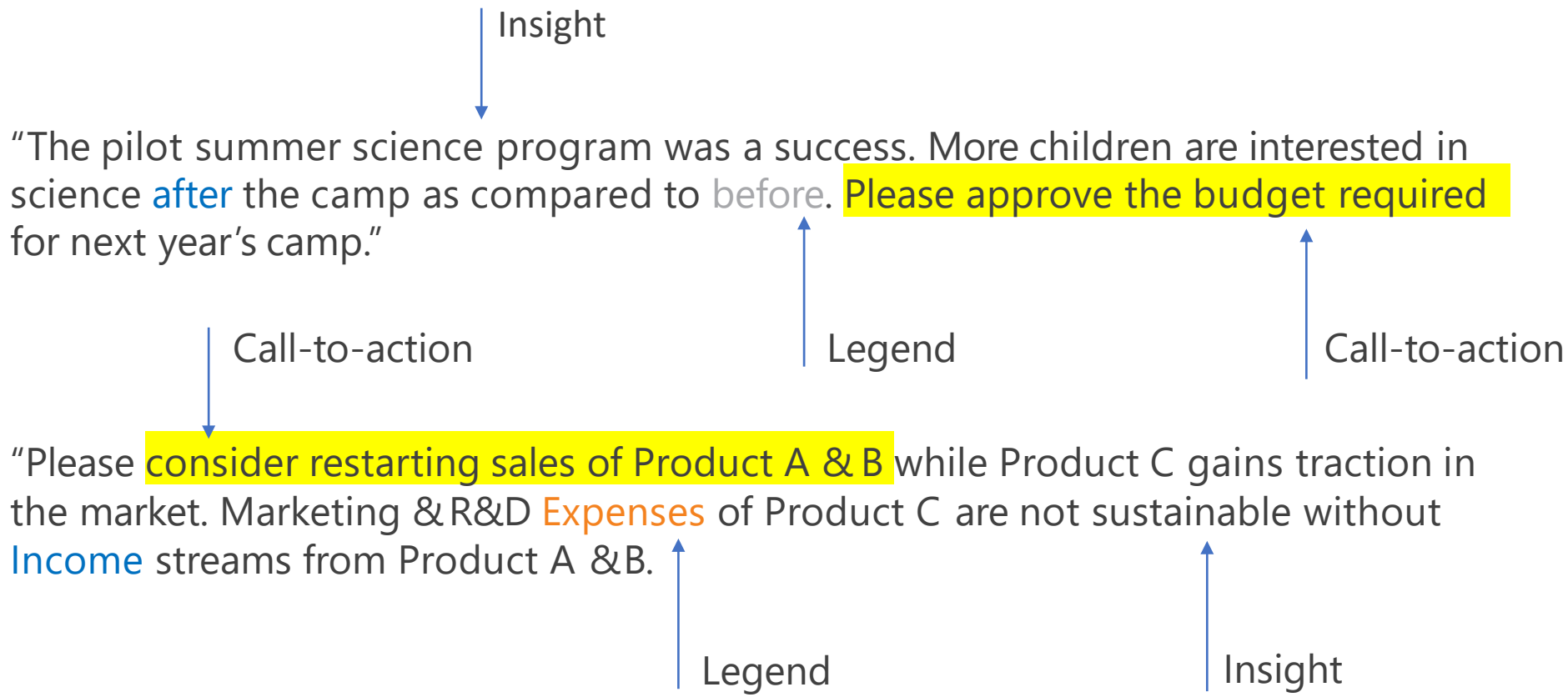
Source: Storytelling with Data pg. 5

"I want the audience to know the camp was a success"



Hybrid Titles

Insight, Call-to-Action & Legend all rolled into one.



What exactly is a call-to-action?

Don't be afraid to start the conversation with these words below.

Prompting action

Here are some action words to help act as thought starters as you determine what you are asking of your audience:

accept | agree | begin | believe | change | collaborate | commence | create | defend | desire | differentiate | do | empathize | empower | encourage | engage | establish | examine | facilitate | familiarize | form | implement | include | influence | invest | invigorate | know | learn | like | persuade | plan | promote | pursue | recommend | receive | remember | report | respond | secure | support | simplify | start | try | understand | validate

Sample phrases using action words

Action words help by telling your audience what you need from them

1. "I want to **persuade** my boss to buy a coffee machine for the office"
2. "Please **empower** your mid-level managers to make their own decisions"
3. "I want my audience to **differentiate** between good & bad storytelling"
4. "**Recommend** a new strategic direction to your Chief Strategy Officer"
5. "I want my audience to **pursue** a healthier lifestyle"

Example – What do I want to tell the audience?

Current situation report & proposed solutions

1. What do I want to show the audience?

Change-over-time for both income & expenses / Absolute value of losses incurred at current date

2. What is the context?

Losses incurred is the result of our change in strategy. We stopped selling Products A & B because we wanted to focus on Product C. It was anticipated that Product C would do well in the market but this has not been the case. Rising R&D / Marketing costs to support Product C not sustainable without revenue from Product A & B.

3. Do I have a call-to-action?

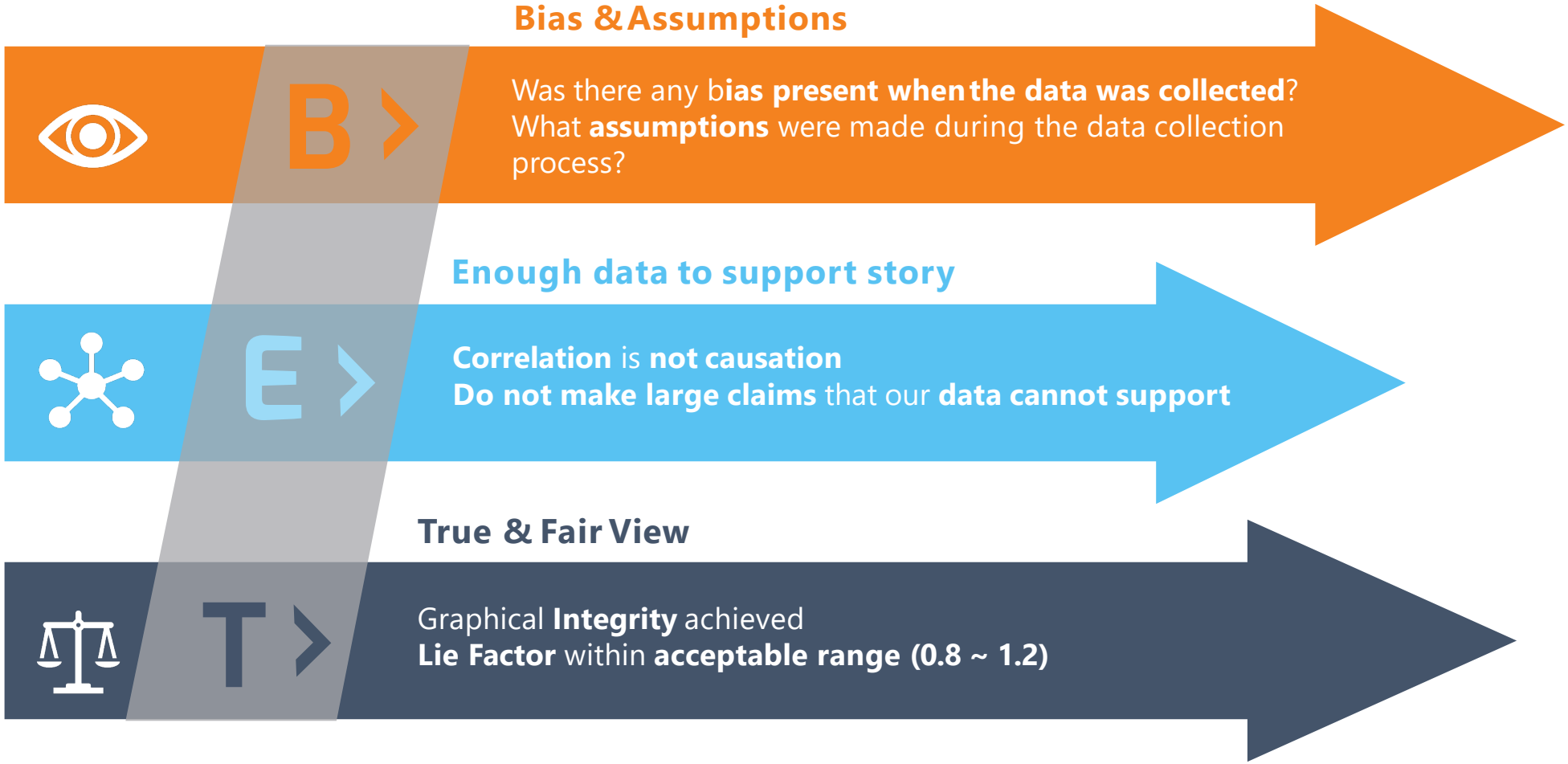
Convince the VP of Strategy to restart sale of Product A & B. Revenue streams will be used to support the further-development of Product C in the market.

How do I tell this story?

Remembering that data is the backbone of our story

Data – Does your data pass the BET test?

Bias & Assumptions, Enough Data, True & Fair View



How – Bias & Assumptions

Check! We aren't often aware of the assumptions of biases we have

Are you asking people inside a KFC restaurant which fried chicken is their favorite? Be mindful of assumptions / bias present in our data.

Person	Restaurant
A	KFC
B	KFC
C	KFC
D	McDonalds
E	KFC

How – Enough data to support our story

Avoid making bold claims we cannot back up with data

Be mindful that the claims we make are supported by our story. Just because our dog barks on days that it rains does not make it a magical dog!

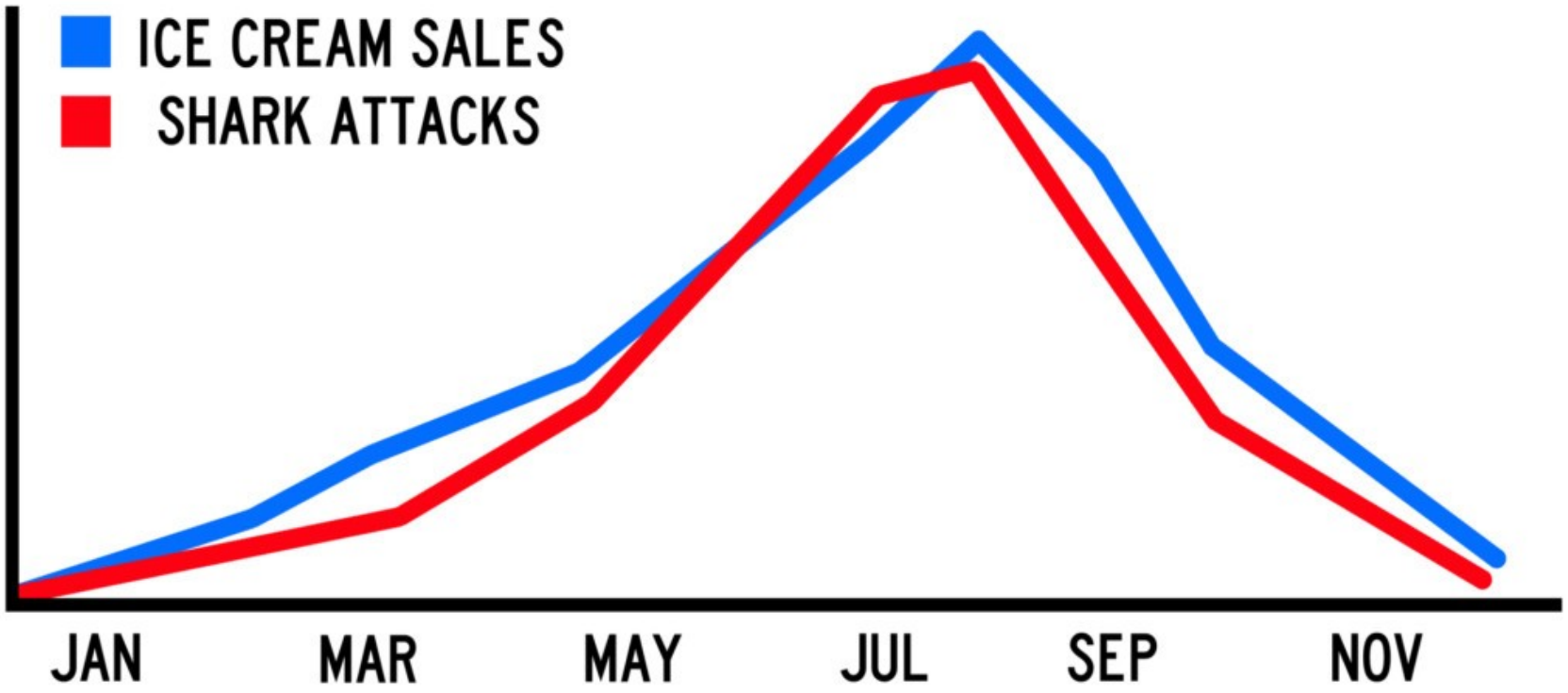
Dog barks in the morning	Rain that evening
Y	Y
N	N
Y	Y
N	N
Y	Y

How – Enough data to support our story

Avoid making bold claims we cannot back up with data

Correlation is not causation

Both ice cream sales and shark attack increase when the weather is hot and sunny, but they are not caused by each other (they are cause by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea.



How – True and Fair View

Be mindful of not misrepresenting your data to the audience

Edward Tufte's Lie Factor

A measure for the True & Fair View

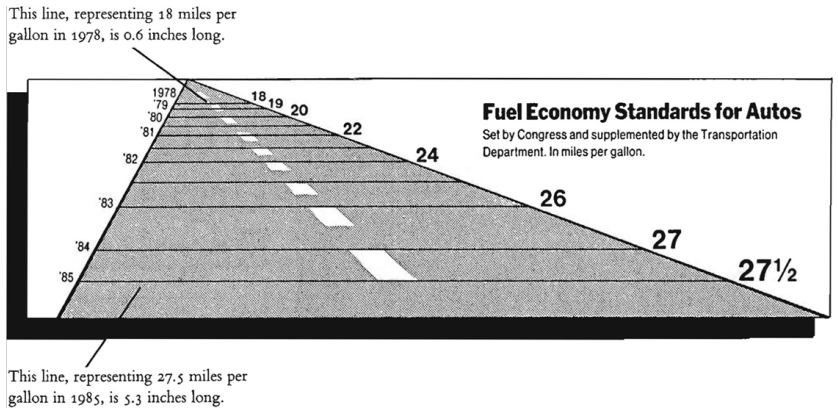
$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

$$\frac{27.5 - 18.0}{18.0} \times 100 = 53\%$$

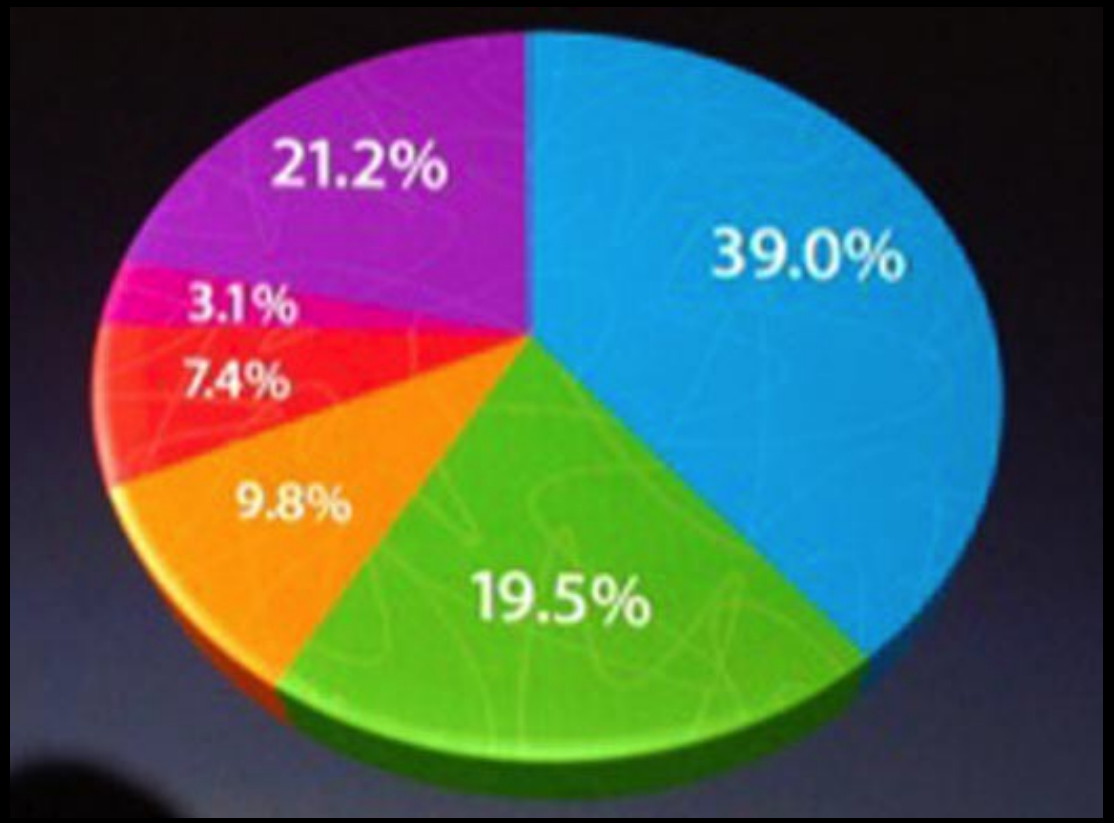
$$\frac{5.3 - 0.6}{0.6} \times 100 = 783\%$$

$$\text{Lie Factor} = \frac{783}{53} = 14.8$$



True & Fair View is about having Graphical Integrity

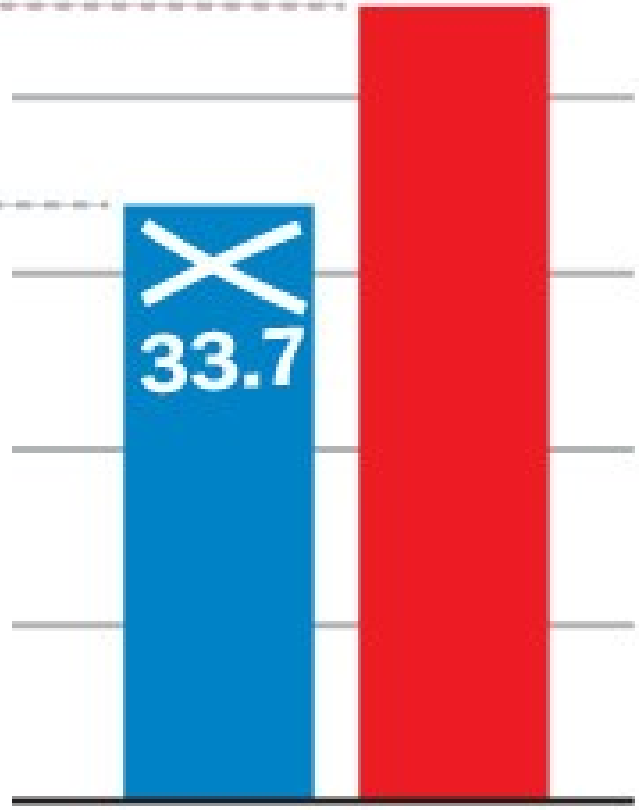
Suspiciously large sales figures of iPhones



Source:Wired.com

Graphical Integrity

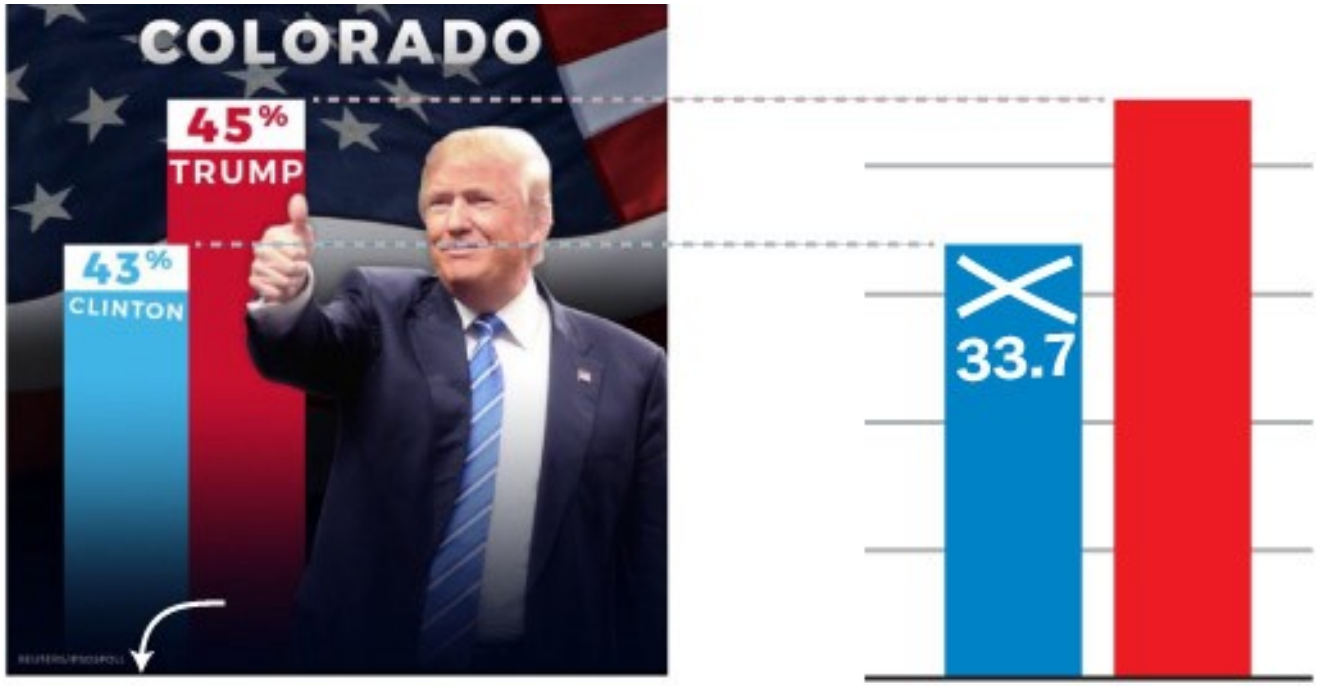
Misrepresentation of Data today



Source: Washington Post

Activity: Calculate the Lie Factor

Misrepresentation of Data today



Source: Washington Post

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

Size of effect in graphic $\frac{6.5 - 4.5}{4.5} \times 100\% = 44\%$

Size of effect in data $\frac{45 - 43}{43} \times 100\% = 4.7\%$

Lie Factor $\frac{44}{4.7} = 9.36$

Anything below 1.0 is understating, and anything above 1.0 is overstating.

Example – How will I tell this Data Story?

We won't look at the underlying data today, but this is a sample BET Checklist

Bias & Assumptions

Assume that we only have three revenue streams (Products A, B & C) / Assume that market failure of Product C can be corrected through further development

Enough data to support my story

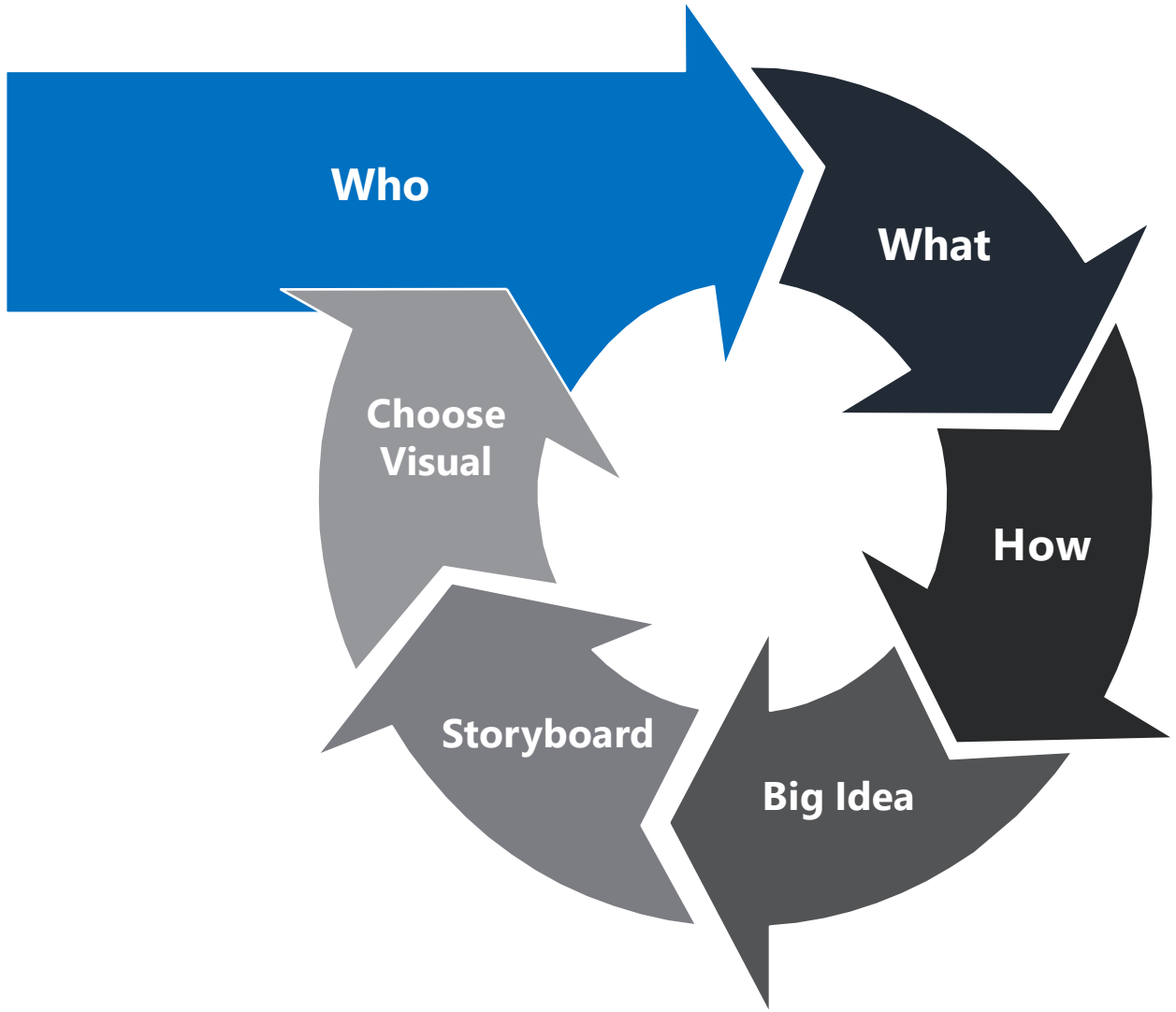
We could supplement our story with the revenue breakdown by product as well as the profit margins & expected sales values for each product / Our underlying analysis shows that this should not be a problem but we could include this information in the appendix because the VP is detail-oriented

True & Fair View

To be calculated after creation of visual

Combining the ‘Who’, ‘What’ & ‘How’ into a Big Idea

This Big Idea is guiding force for the rest of our storytelling process.



Example - Who is our audience?

She is the Vice-President of Strategy

1. What are they like?

Older / Previously from technical background but has been more client facing recently / Detail-oriented / Time-poor

2. Why are they here?

She is looking for a solution to our problem of falling income & rising expenses

3. What keeps them up at night?

She is new to the role and needs a win / Company Shareholders are critical of her ability / Operational losses unsustainable in long-run

4. How can you solve their problem?

Suggest a viable business strategy to curb losses / Back strategy up with appropriate data

5. How can you best reach them?

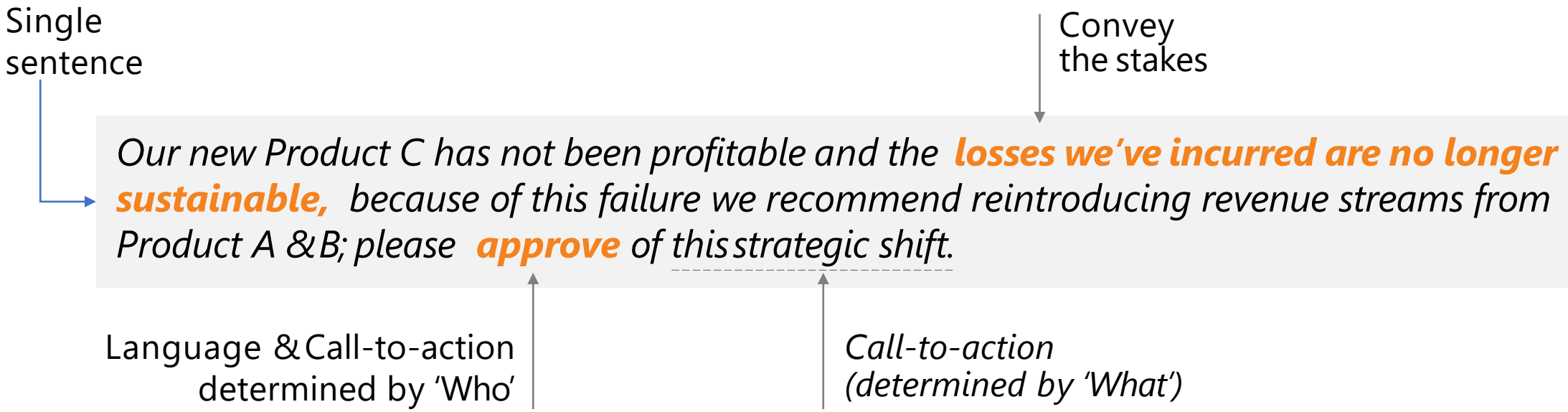
Dislikes lengthy meetings / Tends to rush out last-minute / PowerPoint with Executive Summary & appendix will be good

6. How might they resist?

The proposal must fit within budgetary constraints / needs to be done with readily- available resources

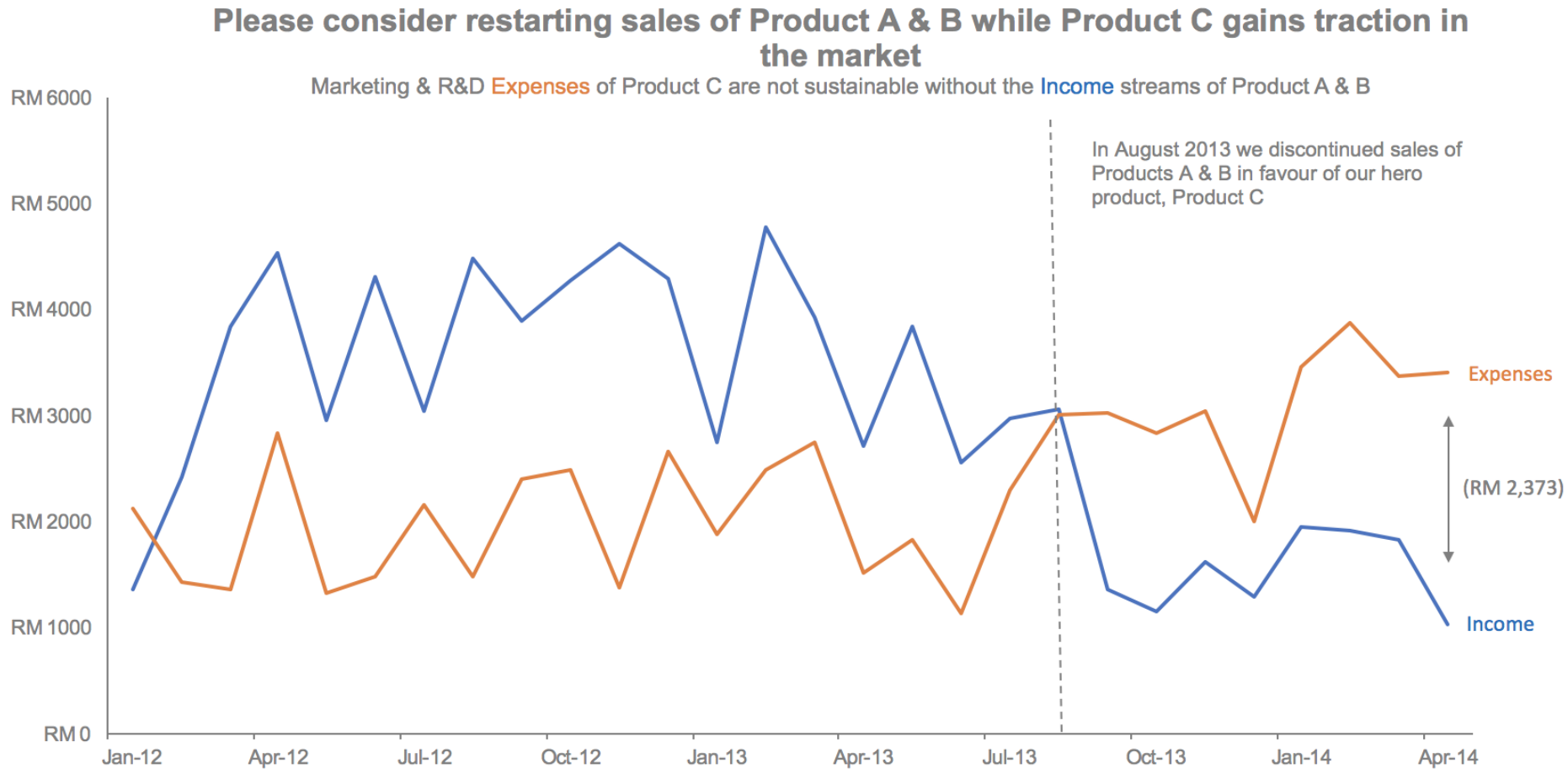
Big Idea – Summarize your findings into a single statement

This high-level statement is similar to an elevator pitch



From Big Idea to Visuals

The new strategy to develop Product C as our hero product has not been successful at capturing the market, because of this failure, we recommend reintroducing revenue streams from Product A & B; please approve of this strategic shift.



Structuring your Narrative

How a Story is told matters

- Different **audiences** react differently.
- **Tailor** storyline and delivery for best results.

Structuring your Narrative

Chronological structure

Grandma is Sick!

Deliver Gift Basket

Grandma's Home

Wolf!

Woodsman Chops

Red Riding Hood V1
Chronological Order

- Story progresses **chronologically**
- Works well when audience cares about the **process**
- Typical approach in management consulting – intuitive to understand

Structuring your Narrative

Reverse Chronological structure

Woodsman Chops

Wolf!

Finds ‘Grandma’

Deliver Gift Basket

Grandma is Sick!

Red Riding Hood V2

Lead with Ending

Trace back her steps from the ending

- Story **begins at the end** and works backwards “Benjamin Button style”
- Works best when audience cares about the **‘So What?’**
- Audience knows **what to look out f**or as we work backwards

  Why did he kill a wolf?

Structuring your Narrative

A quick comparison.

Which version of the story did you prefer?

Analysts generally prefer a **chronological** structure.
Management on the other hand, generally prefers a **reverse-chronological** structure.

Storyboarding & Narrative Structure

In what order should I tell this story?

Two different ways to tell the same story:

Chronological/ Head to Toe

#1	#2	#3
Stop Products A & B	Product C underperforming	Restart sales

Lead with ending/ Toe to Head

#3	#2	#1
Restart sales	Product C underperforming	Stop Products A & B

Chronological Example

How did I arrive at this recommendation?

Chronological/ Head to Toe

#1 Stop Products A & B	#2 Product C underperforming	#3 Restart sales
Focus groups & research shows that children will like Product C because it has broad-appeal to both genders and varied interest groups In August 2013 we stopped sales of Product A & Product B to focus on Product C	Losing revenue streams from Product A & Product B is a huge deal. We cannot sustain our operational losses for long Turns out parents are confused about Product C – should they buy it for boys or for girls? They've never seen this product before	While we still believe in Product C, we need to find ways to supplement our income until sales for Product C kick-off. We will thus recommend the reintroduction of Product A & Product B into the market for the next few months,

Lead with the End Example

Tell them the outcome & let them investigate how you got there

Lead with ending/ Toe to Head

#3 Restart sales	#2 Product C underperforming	#1 Stop Products A &B
Please reintroduce Product A &B to the market as we need the revenue streams to sustain our operational cost.	Costs have been rising to support the marketing & R&D of Product C. We have also lost revenue streams from Product A & B because of a strategic shift on August 2013.	Initial findings suggest that kids would love Product C but on the condition that they did not already own Product A or B. Our company decided to stop sales of A &B to focus on the introduction of Product C.

Which version did you prefer?

Chronological or Leading with the end

Most people prefer Chronological

Why? It follows a natural **three-act structure**

Aristotle's Three Act Structure

Setup, Conflict, Resolution

Aristotle's Three Act Structure

Setup
- What is the background of our characters?
- What is life like during the good times?

Conflict
- An event or problem that the characters have to overcome

Resolution
- The character resolved the event and is forever changed for the better (or worse)

Three Act Structure in the Wild

A Cinderella Story

Setup

Cinderella lives with her stepmother and siblings. She is forced to clean and live in a tower

Conflict

Magic! Her fairy godmother gives her a dress and chariot to attend the ball. She runs away before midnight

Resolution

The prince finally puts the slipper on Cinderella. They get married and we get our happy ending

Three Act Structure in the Wild

Spiderman (2002)

Setup

Peter Parker is a regular high school kid with an interest in photography and a girl named Mary Jane

Conflict

Peters uncle is killed by the villain - Green Goblin. Peter hunts down this villain as Spiderman

Resolution

Peter mourns the death of his uncle and his best friend's father. He carries on as Spiderman to protect the city

Looking at our Chronological Story through the Three Acts

Setup – Conflict - Resolution

#1
Stop Products A & B

Focus groups & research shows that children will like Product C because it has broad-appeal to both genders and varied interest groups.

In August 2013 we stopped sales of Product A & Product B to focus on Product C

#2
Product C underperforming

Losing revenue streams from Product A & Product B is a huge deal. We cannot sustain our operational losses for long.

Turns out parents are confused about Product C – should they buy it for boys or for girls? They've never seen this product before.

#3
Restart sales

While we still believe in Product C, we need to find ways to supplement our income until sales for Product C kick-off.

We will thus recommend the reintroduction of Product A & Product B into the market for the next few months.

Stories stay in the mind long after presentations end

Example - Storyboarding

Remembering our VP of Strategy

Lead with Ending

#3
Restart sales

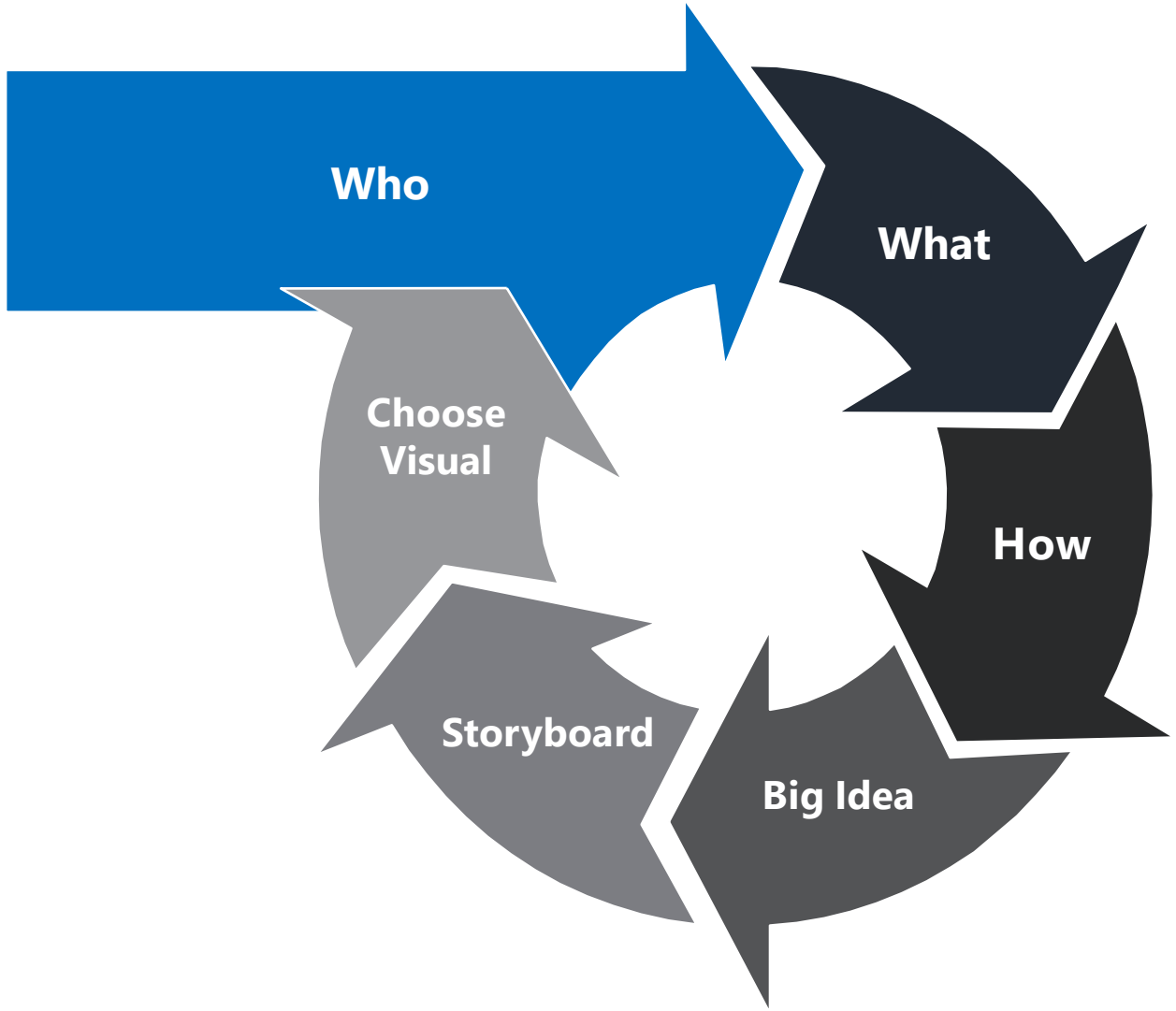
#2
Product C underperforming

#1
Stop Products A & B

1. **Time-poor** individual
2. We have a **working-relationship** & she **trusts** my analysis
3. Information can be added in the **appendix** (detail-oriented)

We are ready to start choosing visuals!

Though it may seem excessive, these steps are **more efficient** than randomly creating visuals



Data Visualization

Visualizing your Data

The right **Visualization** brings **Data to life**

- Cater to specific **data type**
- **Simplicity** is key
- **Interpretability** matters

Photo credit: Towards Data Science

Visualizing your Data

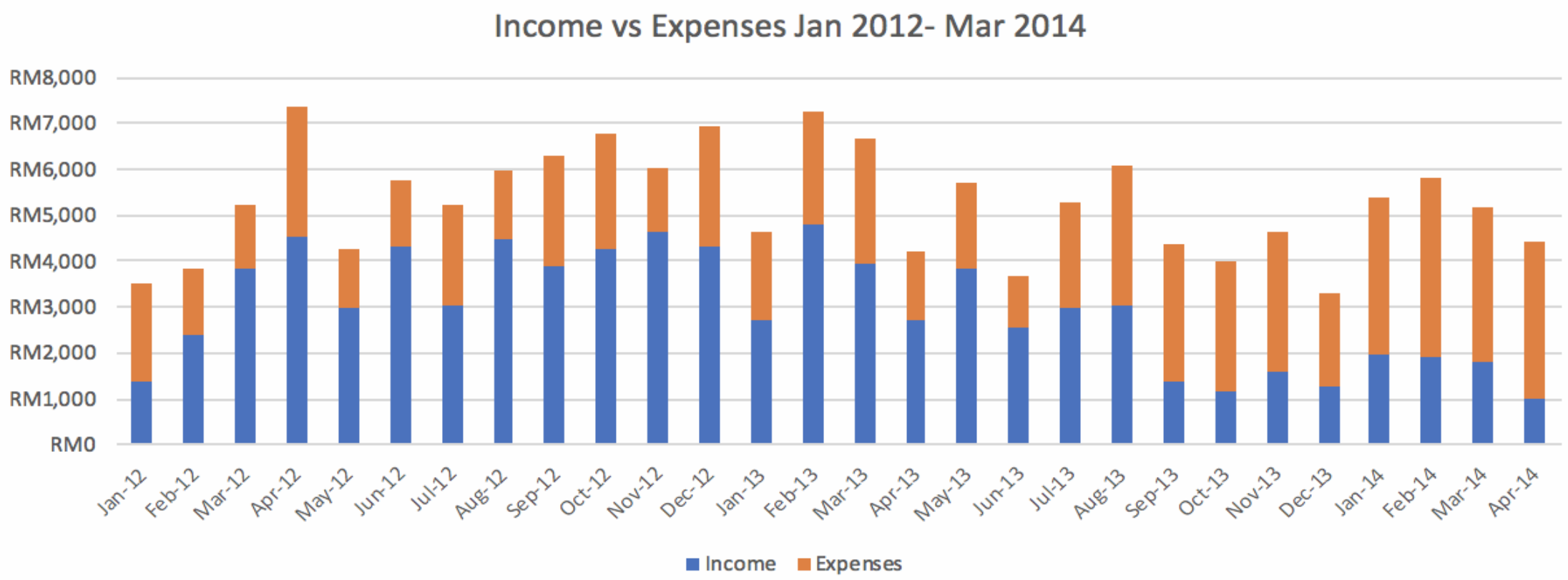
Marketing & Sales scenario.

Photo credit: Complex SQL

- A company sells two products A and B
- In Aug 2013, they discontinue sales of both products in favor of a successor product C
- Analysis performed based on income and expenses data for Jan 2012 – Apr 2014

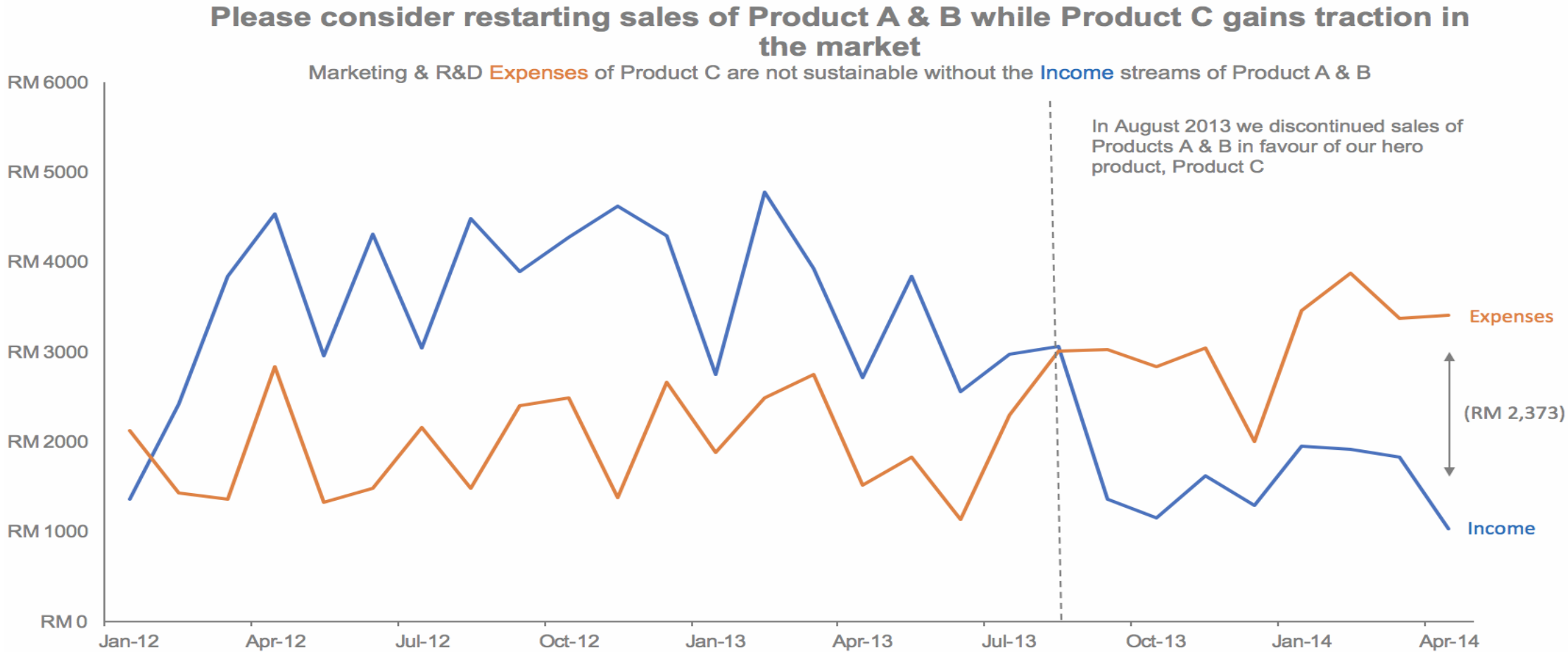
Visualizing your Data

Poor Visualization



Visualizing your Data

Good Visualization



Graphical Perception

(McGill & Cleveland, 1984)

The Visual Decoding of Information Encoded on a Graph

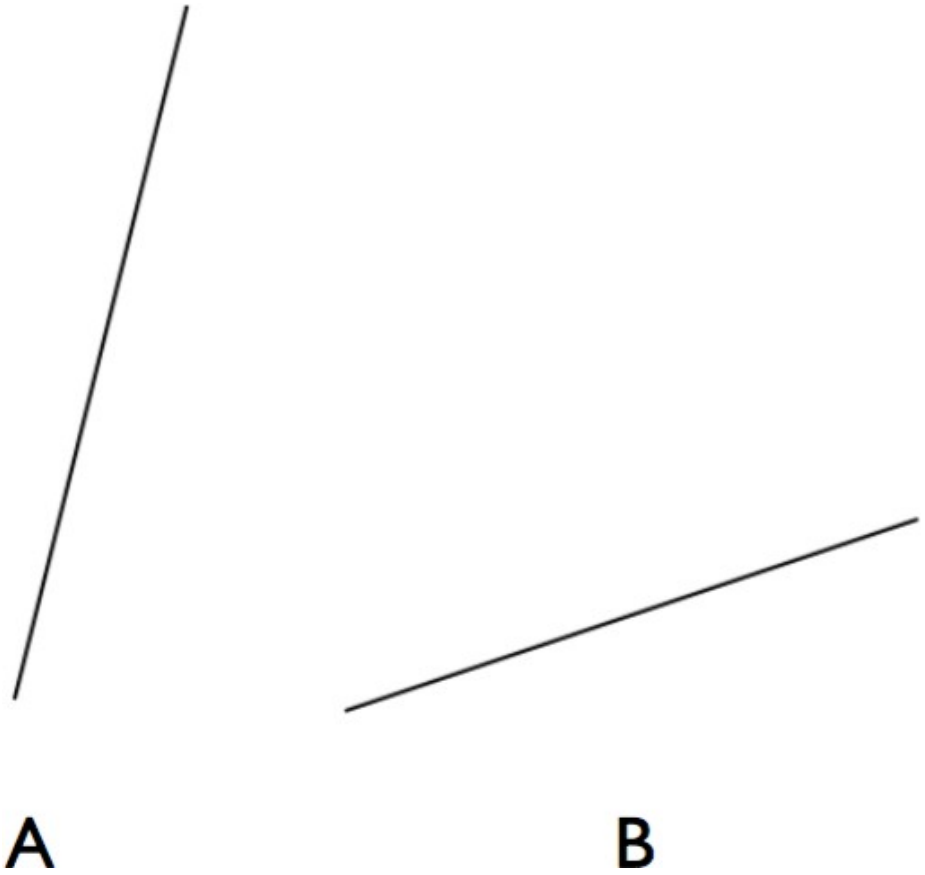
How much longer is B compared to A?

Graphical Perception: Elementary Perceptual Tasks

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

How much steeper is A compared to B?

Graphical Perception: Elementary Perceptual Tasks

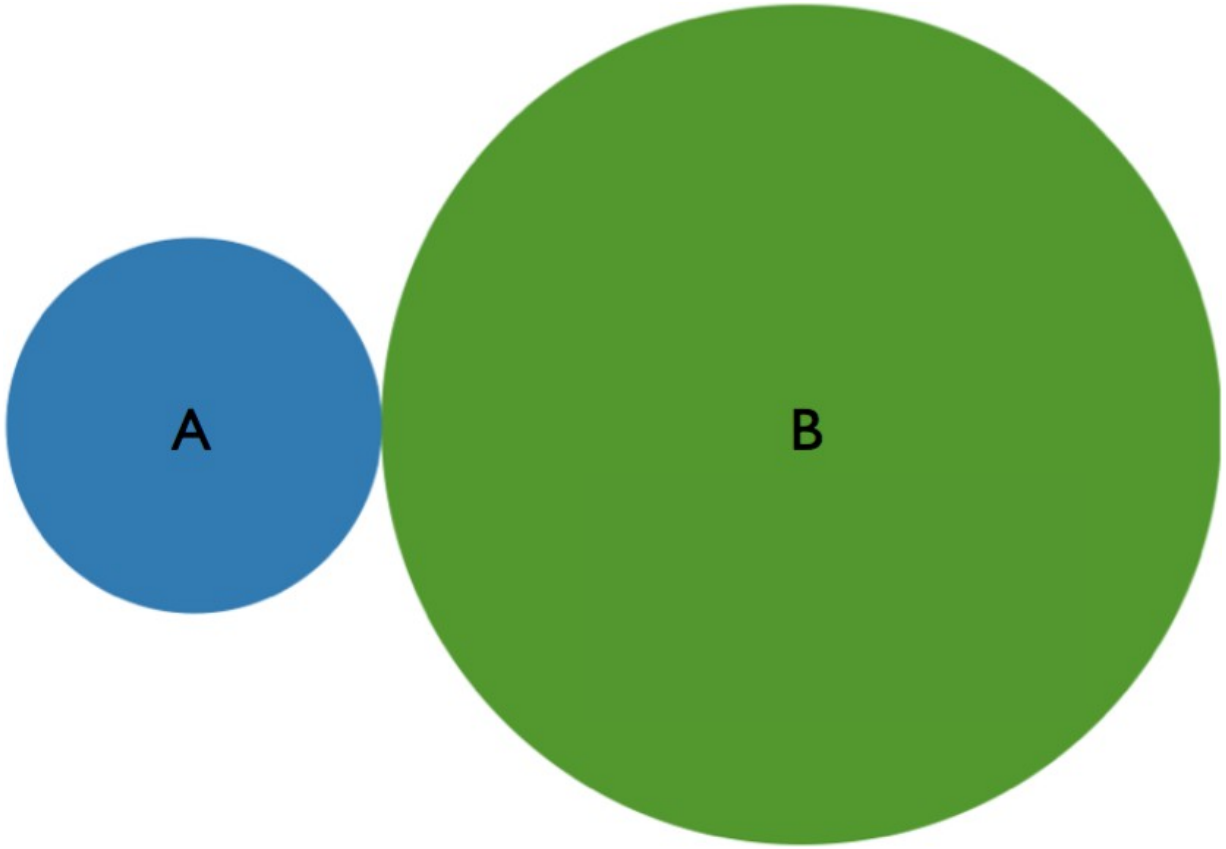


6x

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

How much larger is B compared to A?

Graphical Perception: Elementary Perceptual Tasks



10x

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

How much darker is B compared to A?

Graphical Perception: Elementary Perceptual Tasks

A

B

2x

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

How much bigger in value is B compared to A?

Graphical Perception: Elementary Perceptual Tasks

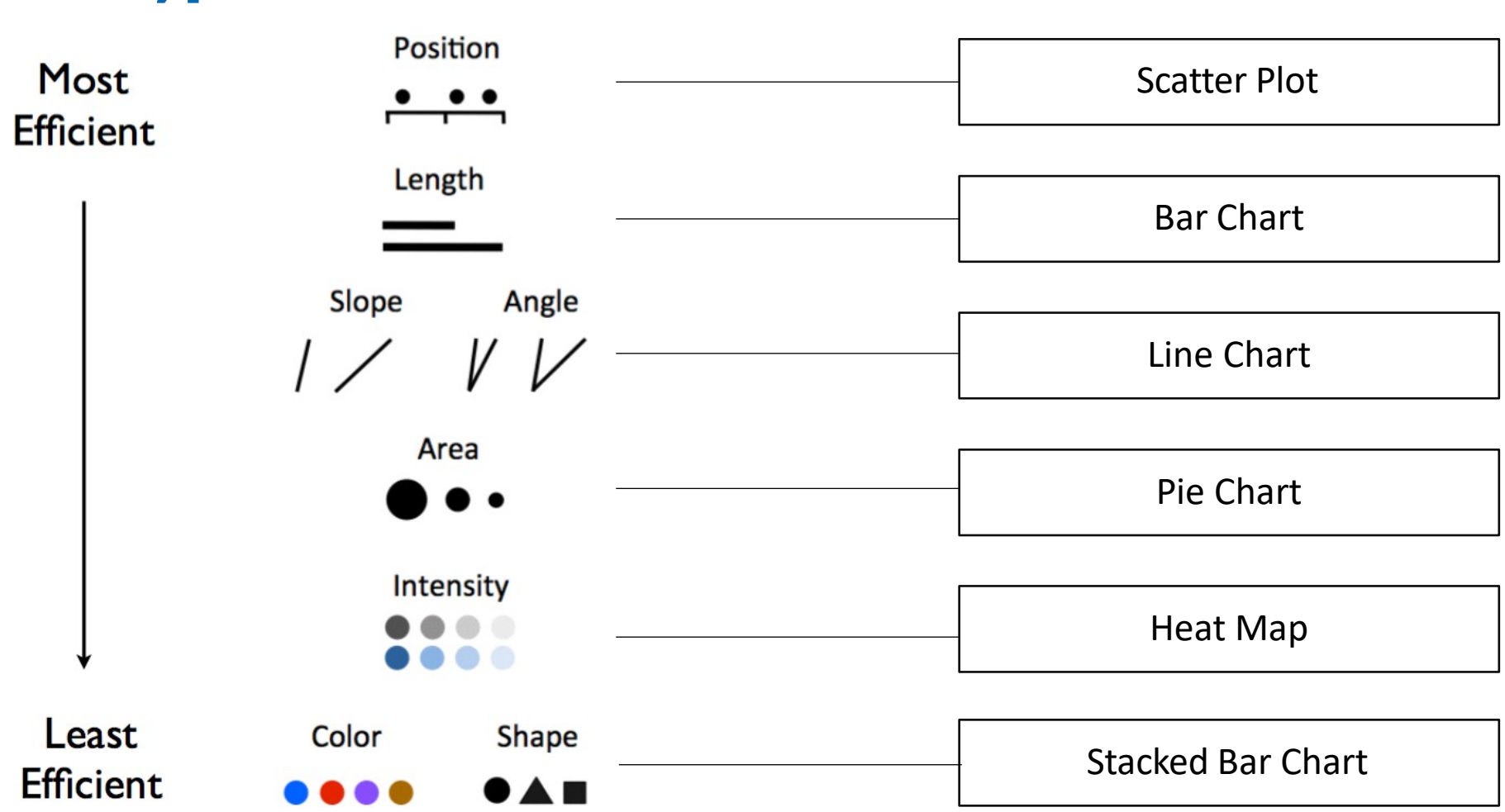
A

B

4x

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

Which perceptual tasks are used by common chart types? McGill & Cleveland (1964)



Visual Perception

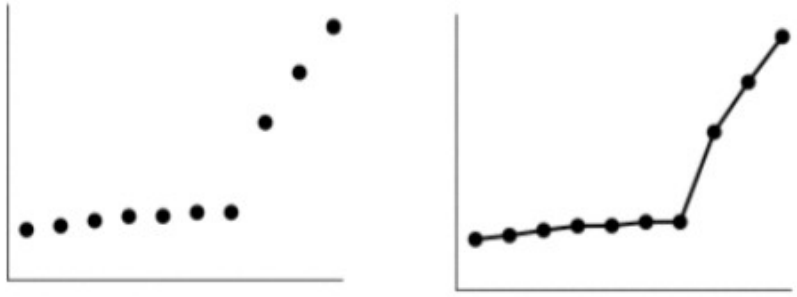
People are more inclined to perceive certain
visual cues (variables) better than **non-visual cues.**

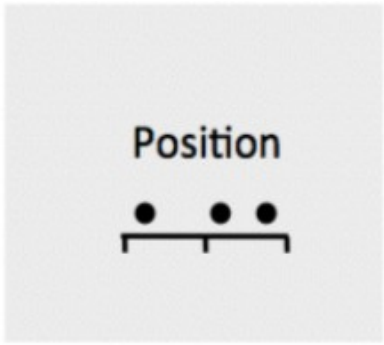
MORE ACCURATE — LESS ACCURATE

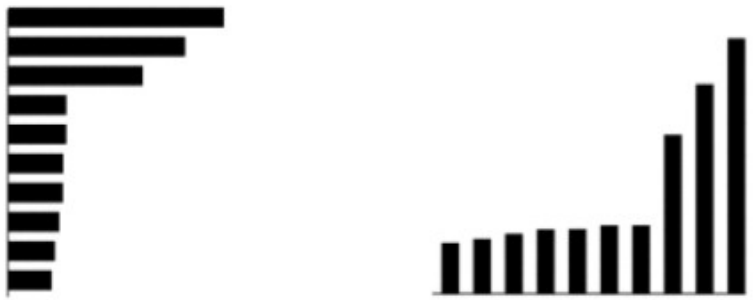
Position | Length | Angle | Direction | Area | Volume | Saturation | Hue

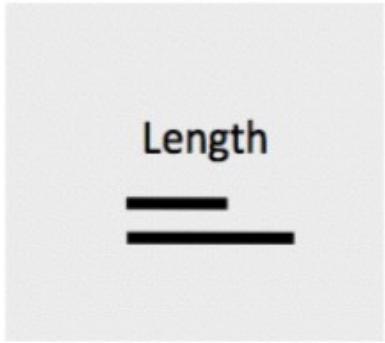
Dot Plots & Bar Charts are very effective

They use position and length to convey information



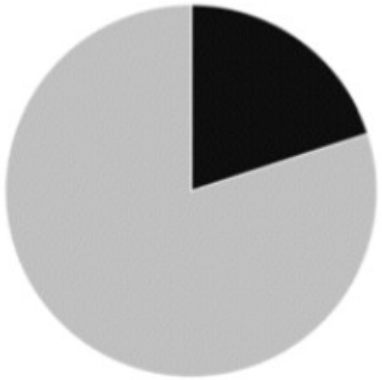


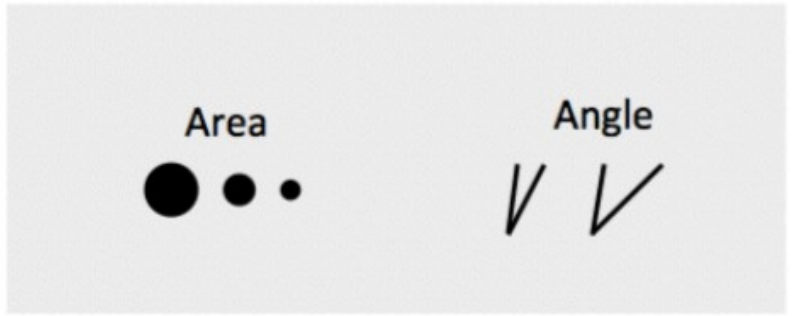


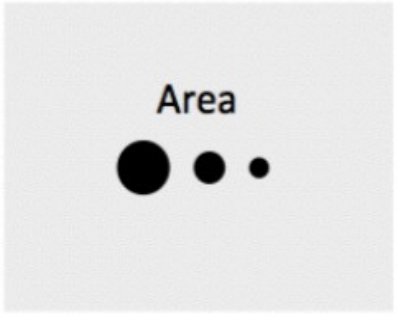


Pie Charts & Donut Charts are less effective

The human-eye cannot perceive 2D-area and angle that well







Bar Charts – Compare between categories

Very common charts – easy to interpret

Rule of Thumb: **Four multi-series maximum**

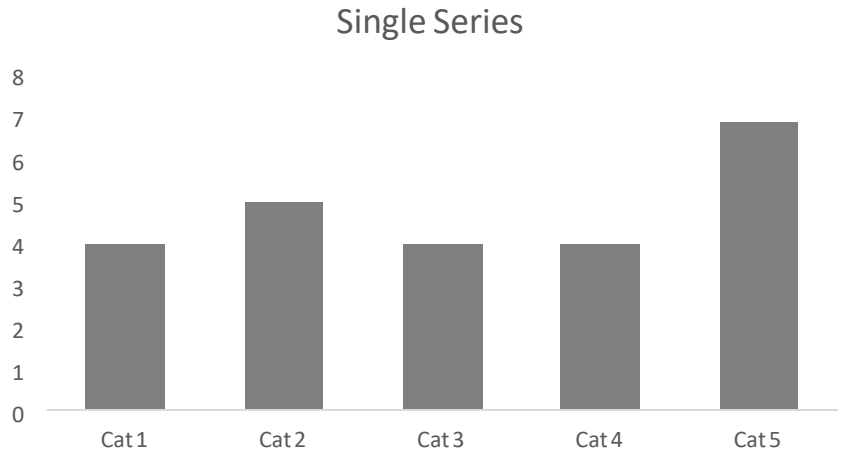
- Multiple series can quickly overwhelm the audience

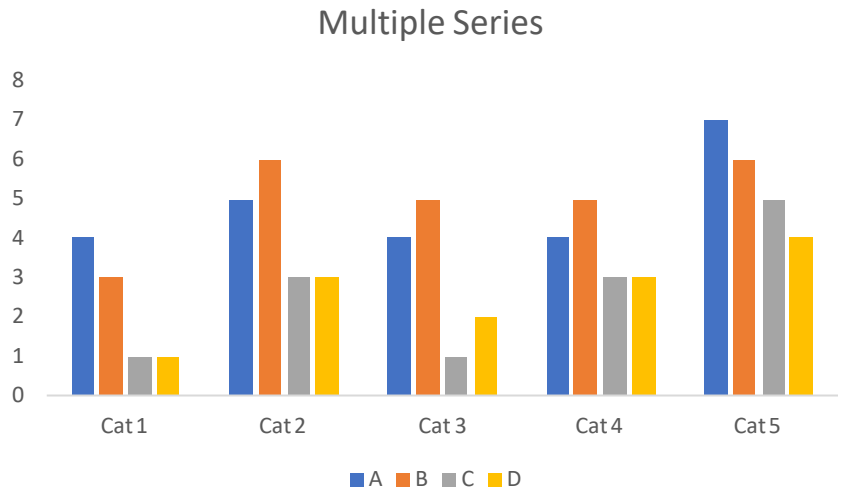
Visual grouping happens as a result of the spacing in bar charts having more than one data series

- **Spacing between bars** for same category must be smaller than between categories.
- **Spacing between categories** should be approx. 0.5 of bar width.

Categories must be ordered in a logical manner such as:

- **Alphabetical** order.
- By **relative importance.**





Horizontal Bar Charts

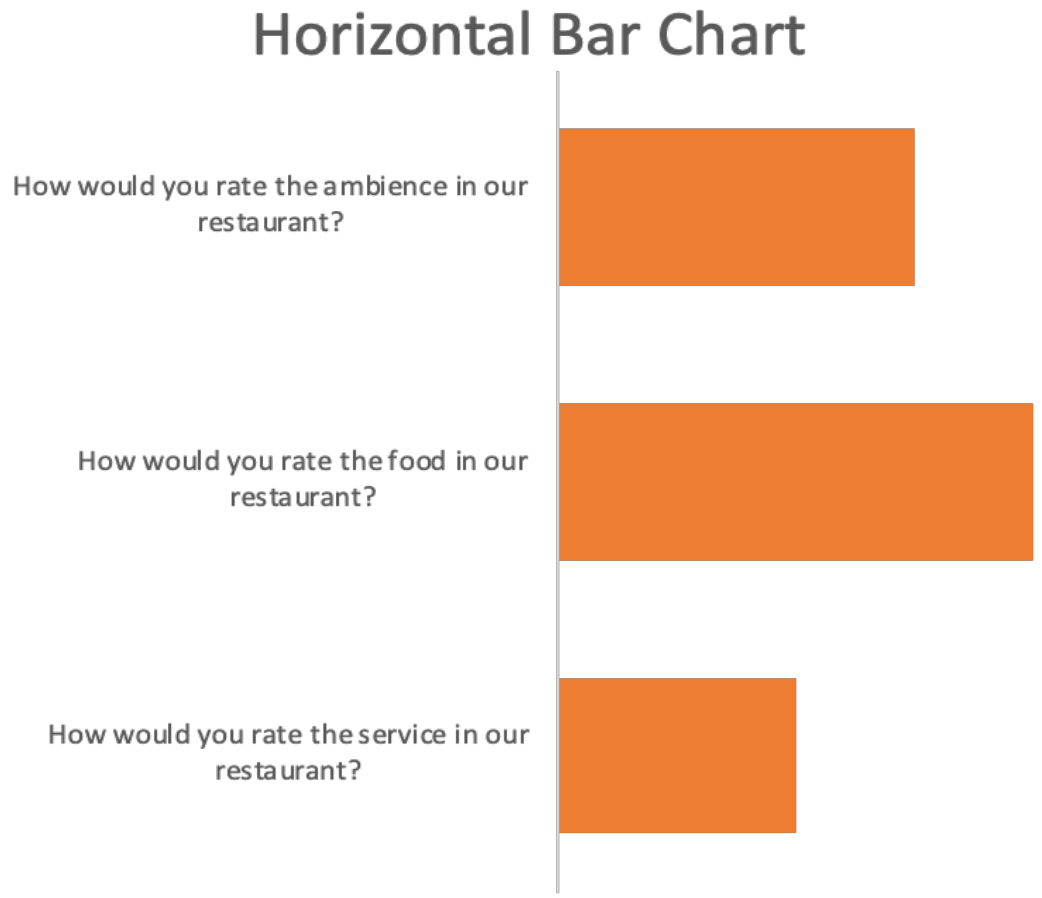
Horizontal Orientation is great for surveys

Useful for long category names as the text is written from left to right (natural for most audiences to read) making the graph more legible

Helps to process the displayed content easily because we read the category names before the actual data (visual hierarchy)

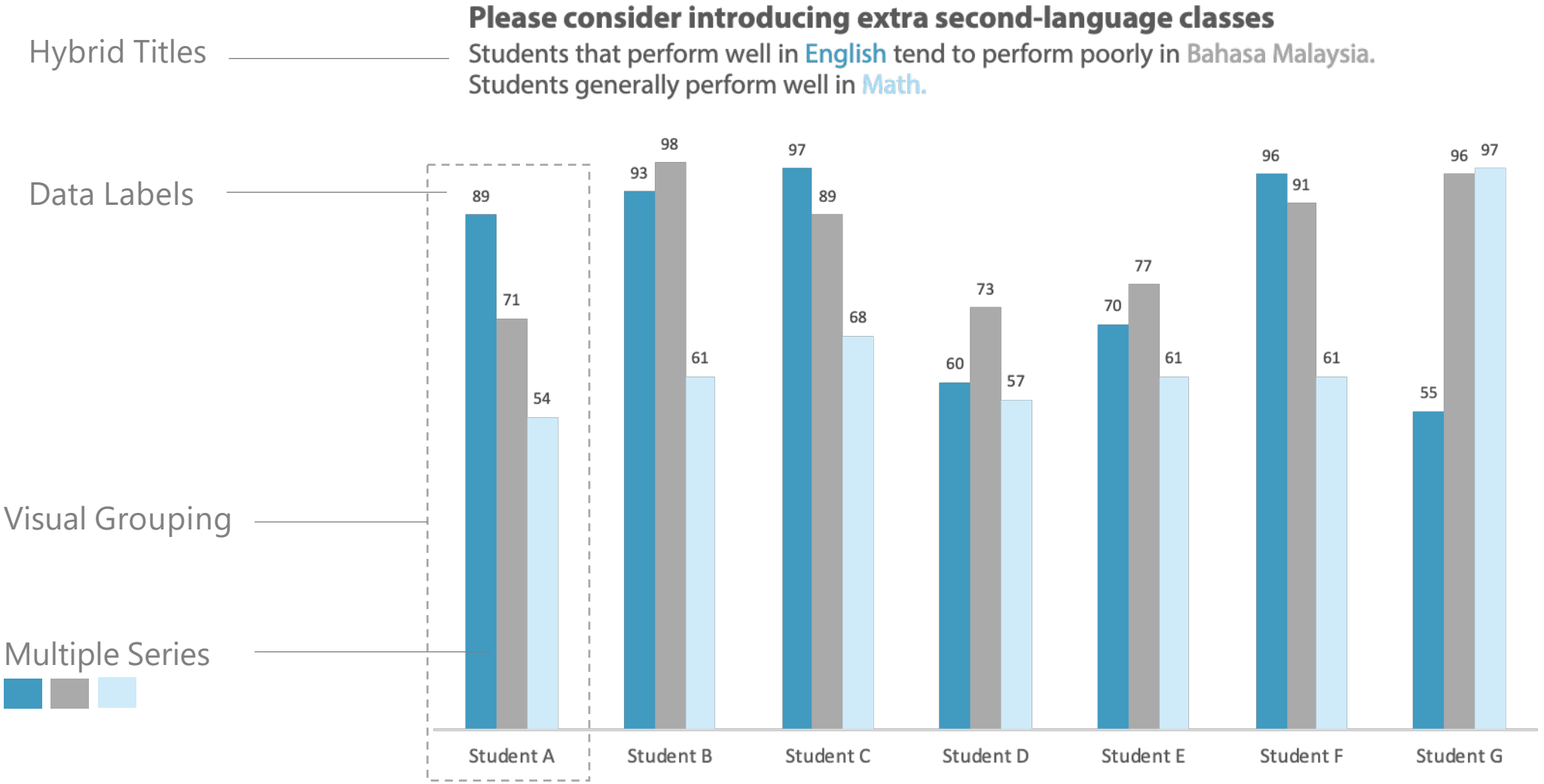
- In contrast, vertical bar charts forces our eyes to move several times between data and category names

Adding data labels provide better accuracy



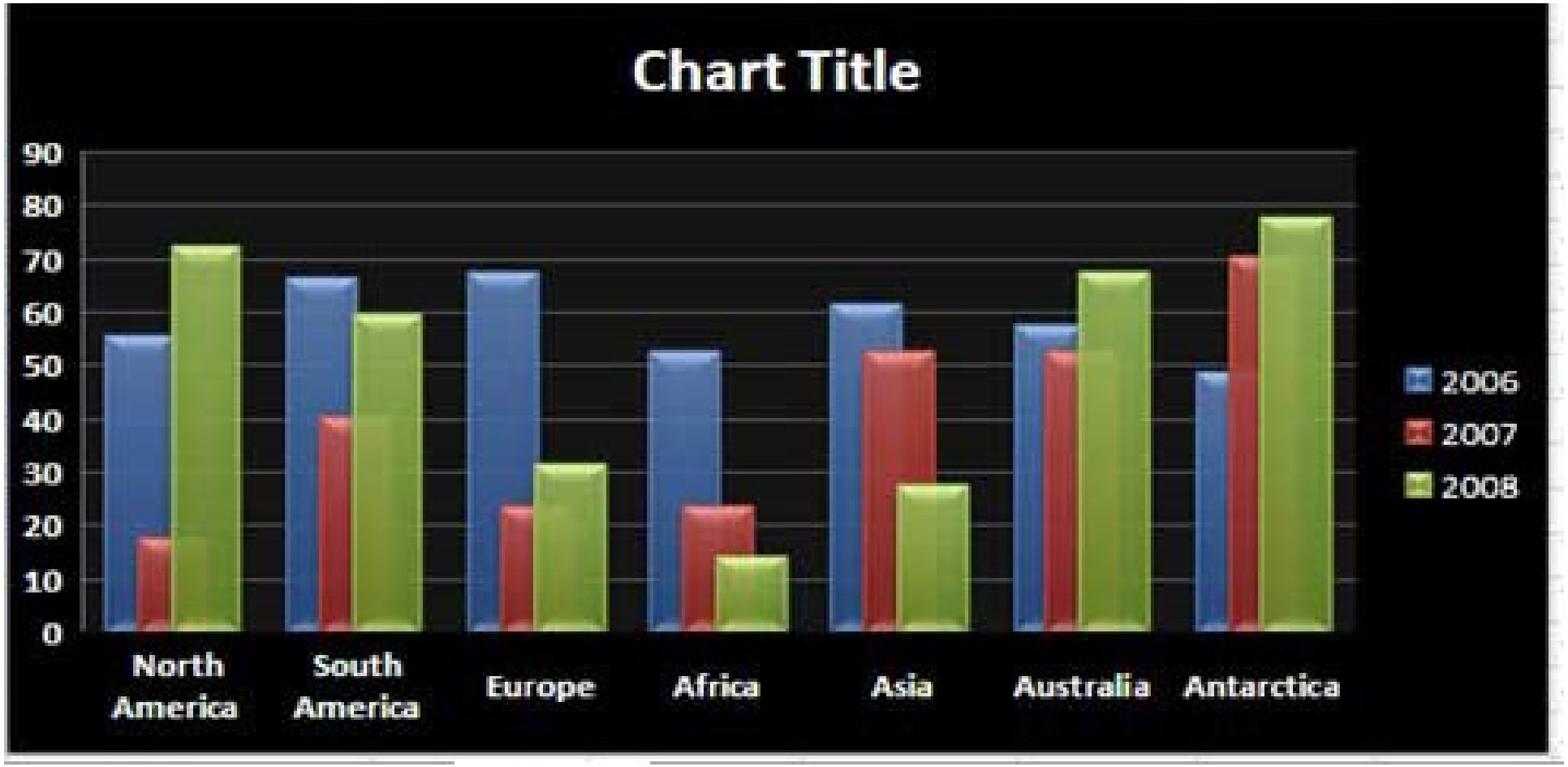
Bar Chart – Best Practice

Can you spot the mistake?



Bar Charts – Bad Example

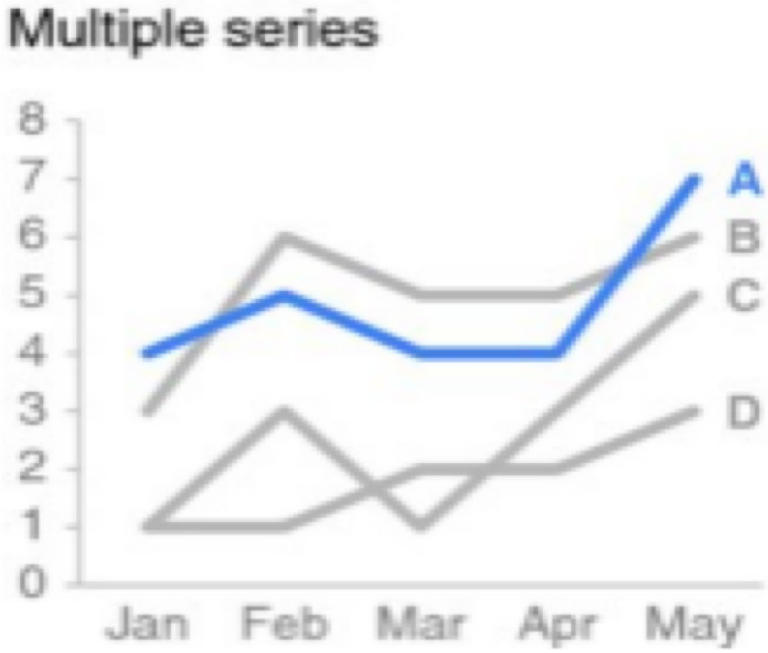
Point out all the flaws on this chart



Line Graphs – Single Variable Change Over Time

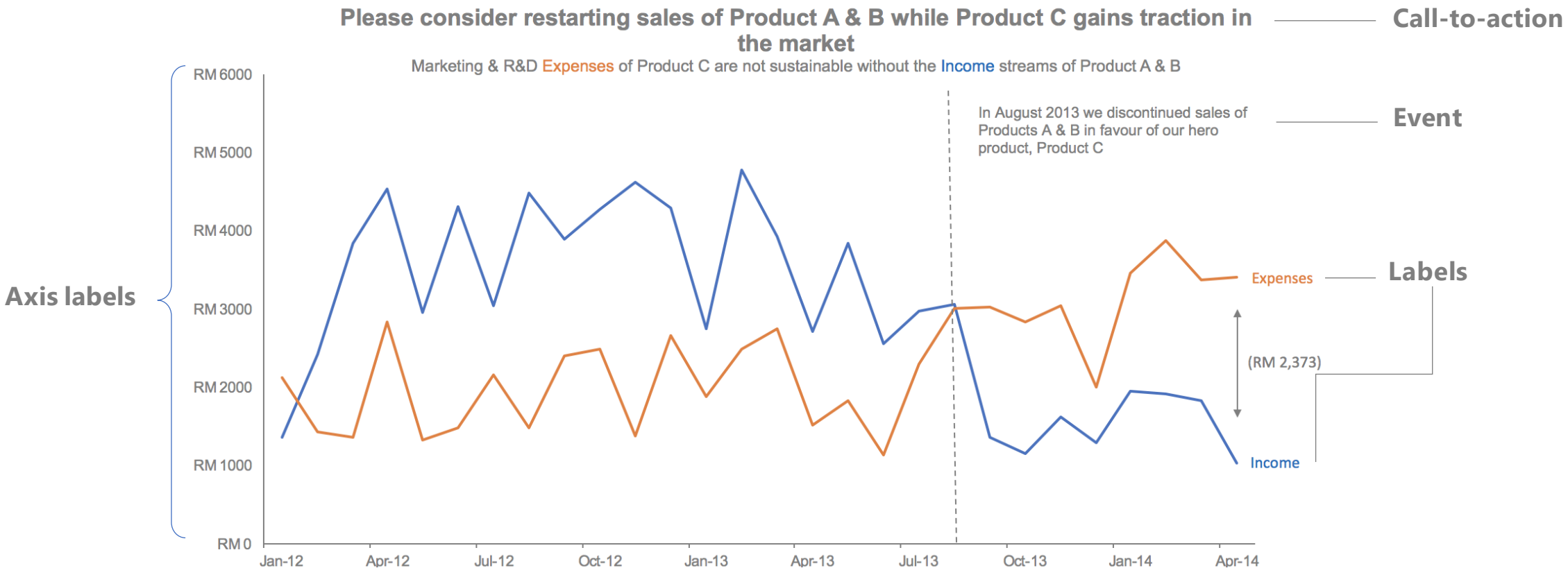
Good for time series data

- The line graph can show a single series of data, two series of data, or even multiple series
- When using multiple series its best to highlight the series that you want your audience to notice
- Used for plotting continuous data
  - Points are physically connected via the line
  - Implies a connection between the points (continuity)



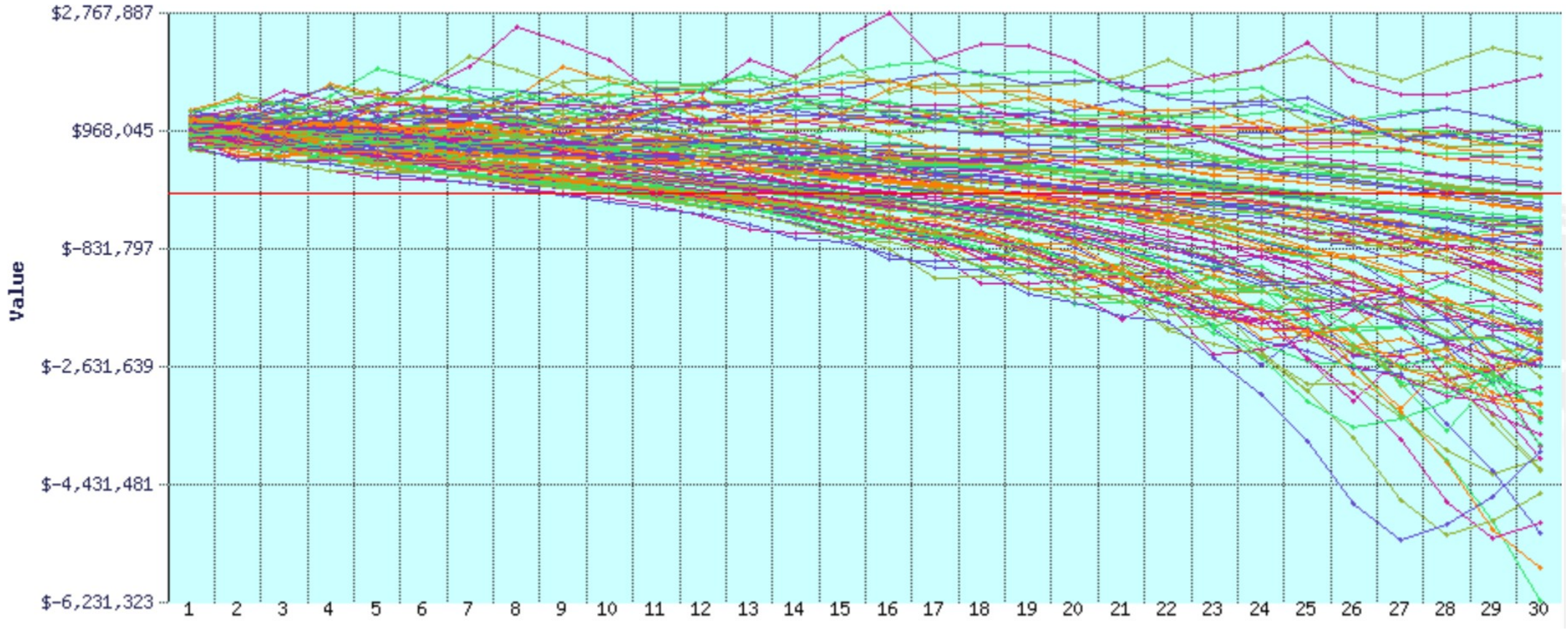
Line Graphs – Best Practice

Annotation of significant events are nice to have for context



Line Graphs – Bad Example

Don’t make a spaghetti graph



Stacked Bar Chart

Proportion to whole & comparison between categories

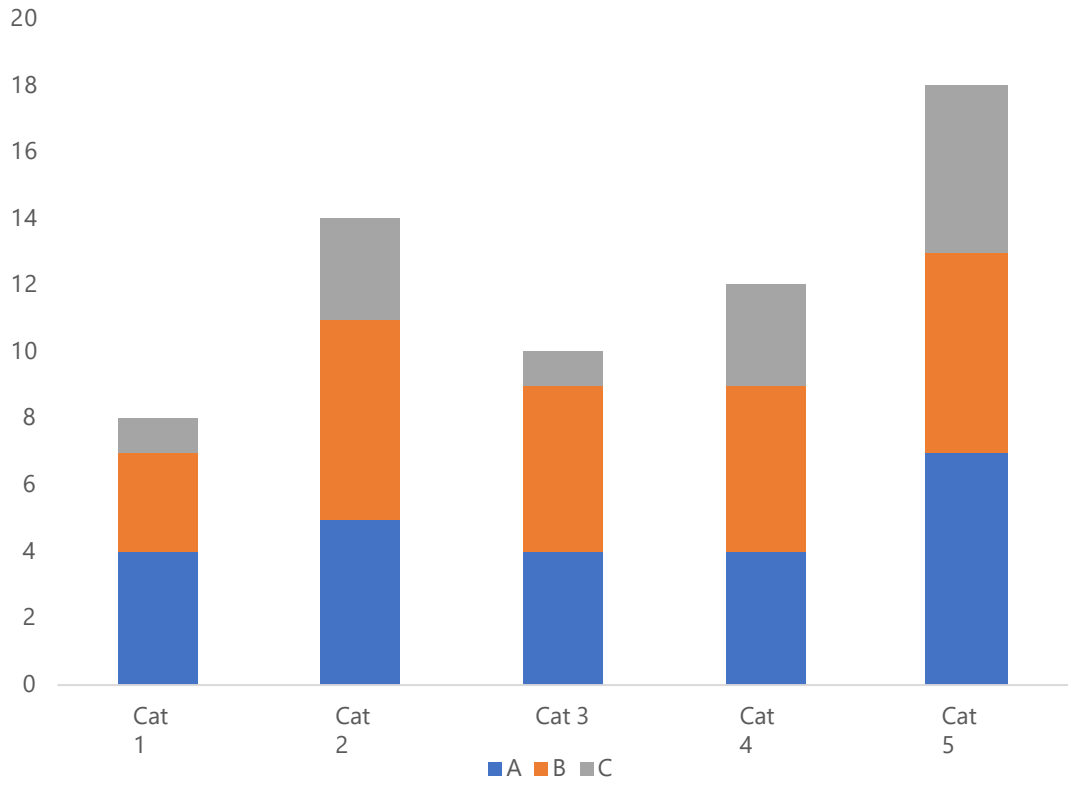
Help **compare totals across categories** and also see the **subcomponent proportions** within a given category

- This can quickly become visually overwhelming

Hard to **compare the subcomponents across** the **various categories** beyond the bottom series

- We no longer have a consistent baseline to use to compare.

Many sub-categories can be difficult to compare



Horizontal Stacked Bar Chart

Normalize to 100% for easier comparison

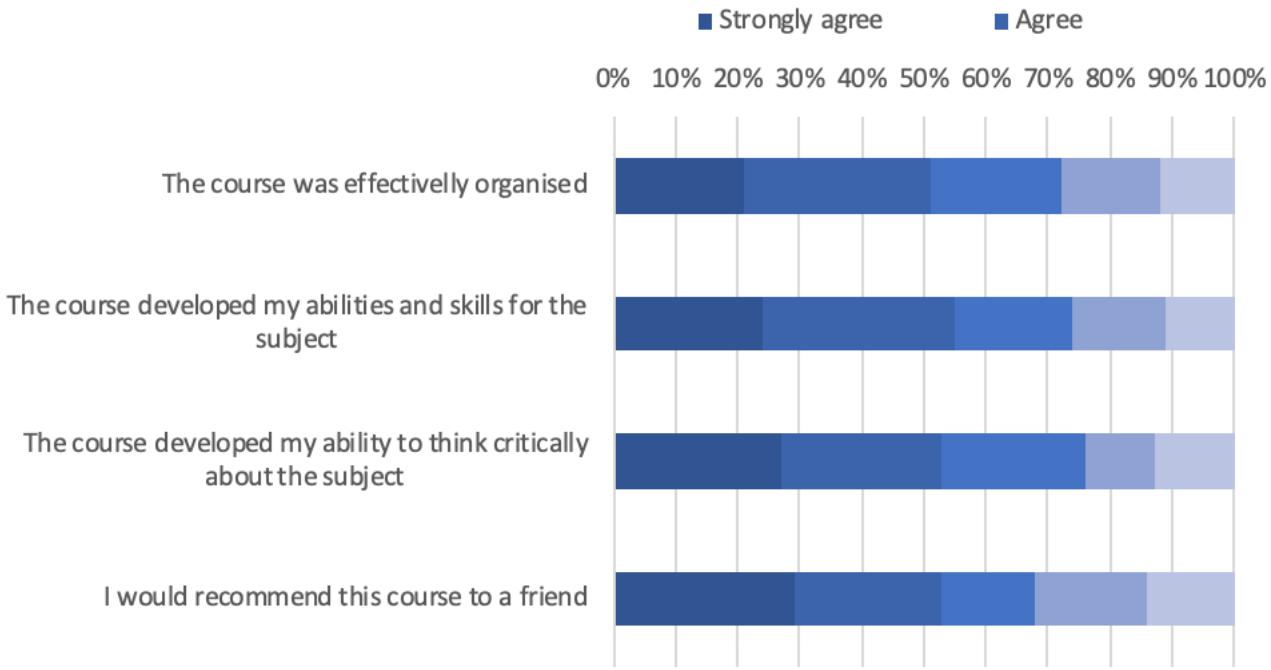
Show **totals for categories** summed to **100%**

- Good for comparison if there are **uneven number of responses** for each question

More useful for visualising **proportions to a whole** on a because we get **a consistent baseline** on both the far left and the far right

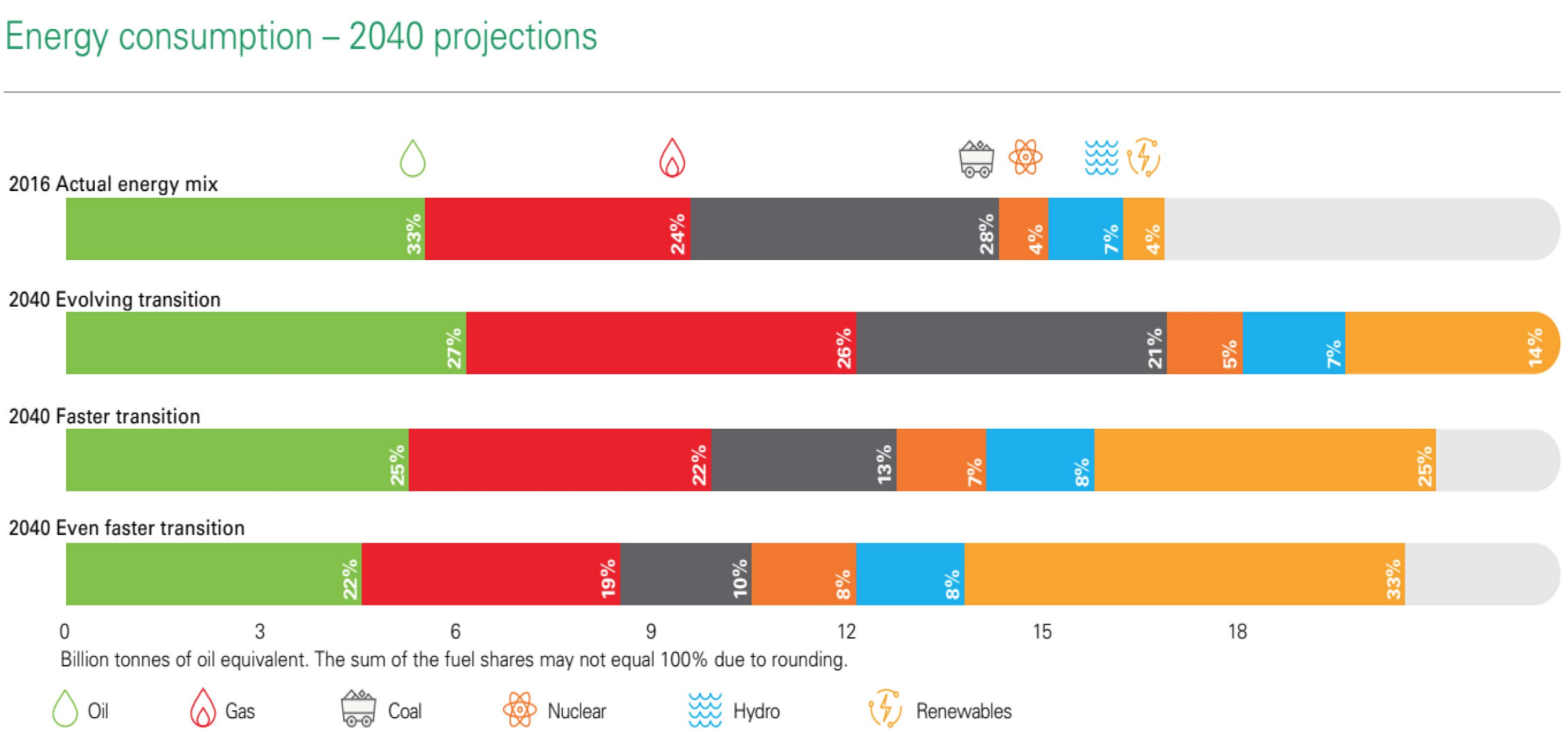
- Easy comparison of the **left-most pieces as well as the right-most pieces**

Visualisation of survey data collected along a Likert scale (a scale commonly used in surveys that ranges from Strongly Disagree to Strongly Agree)



Stacked Horizontal Bar Chart – Bad Example

Cognitive Overload



Pie Charts

Proportions done poorly – Listen to the experts

3D-Effect

Labels

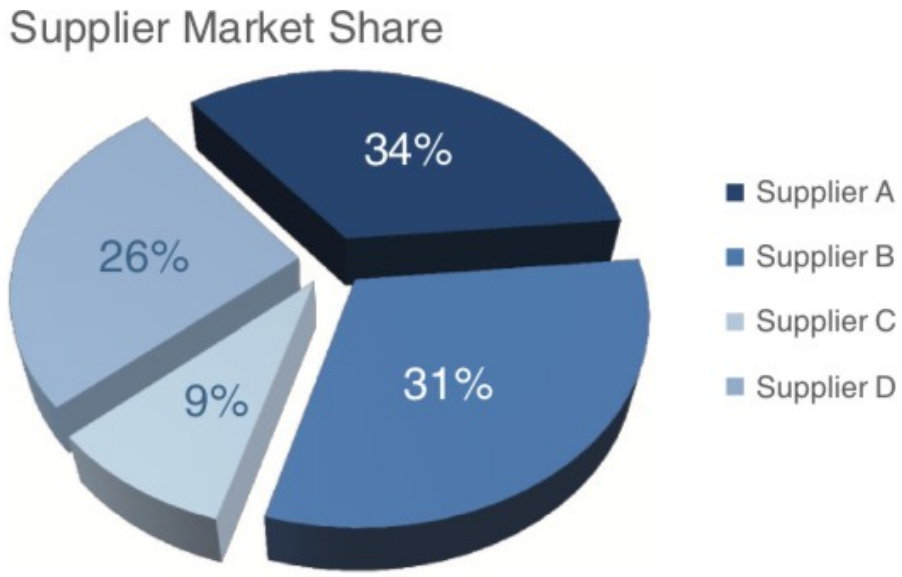


FIGURE 2.22 Pie chart with labeled segments

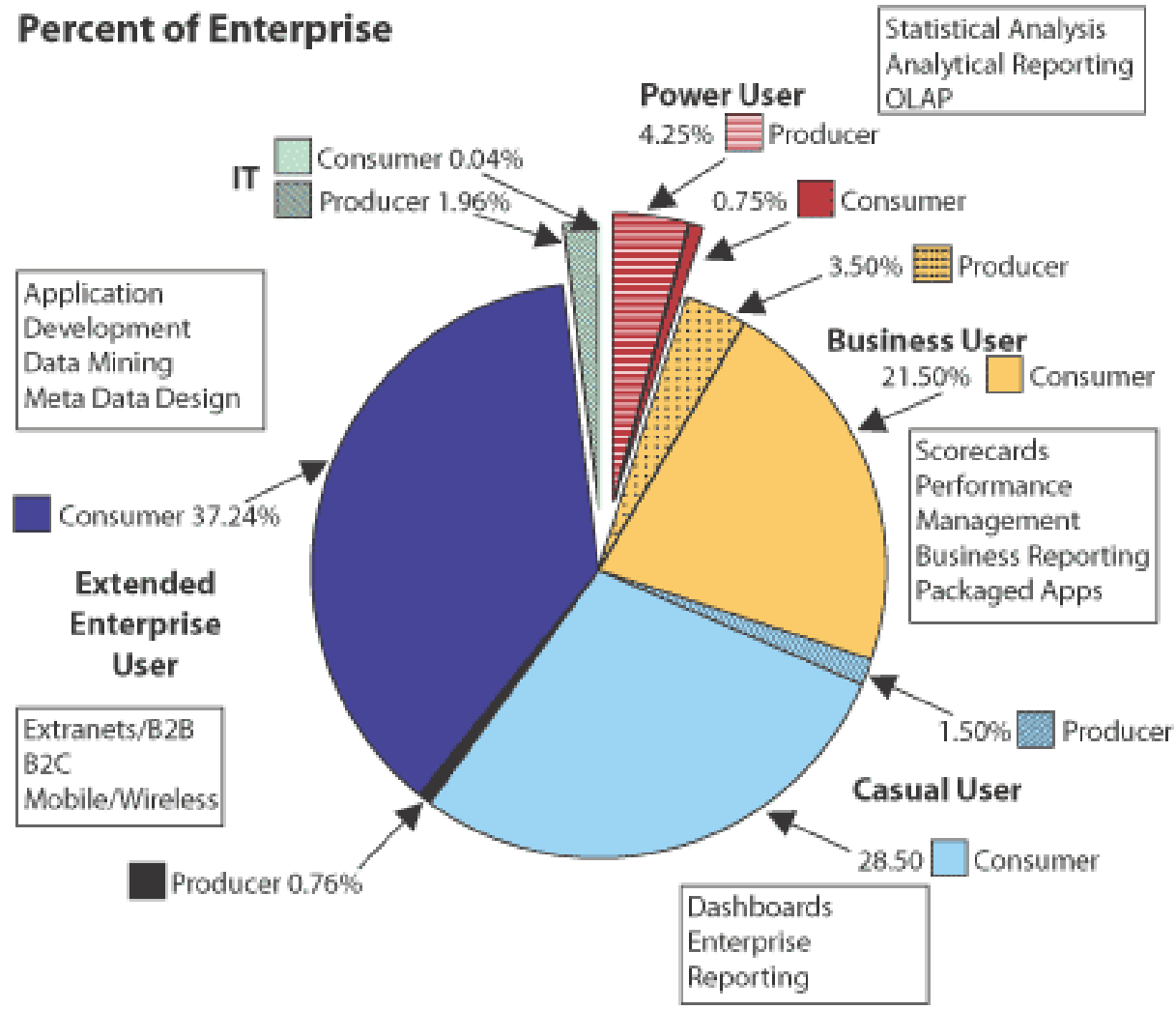
"Save the pies for dessert"
Stephen Few

"Death to pie charts"
Cole Nussbaumer

"The only thing worse than a pie chart is several of them"
Edward Tufte

Pie Chart – Bad Example

Too much!

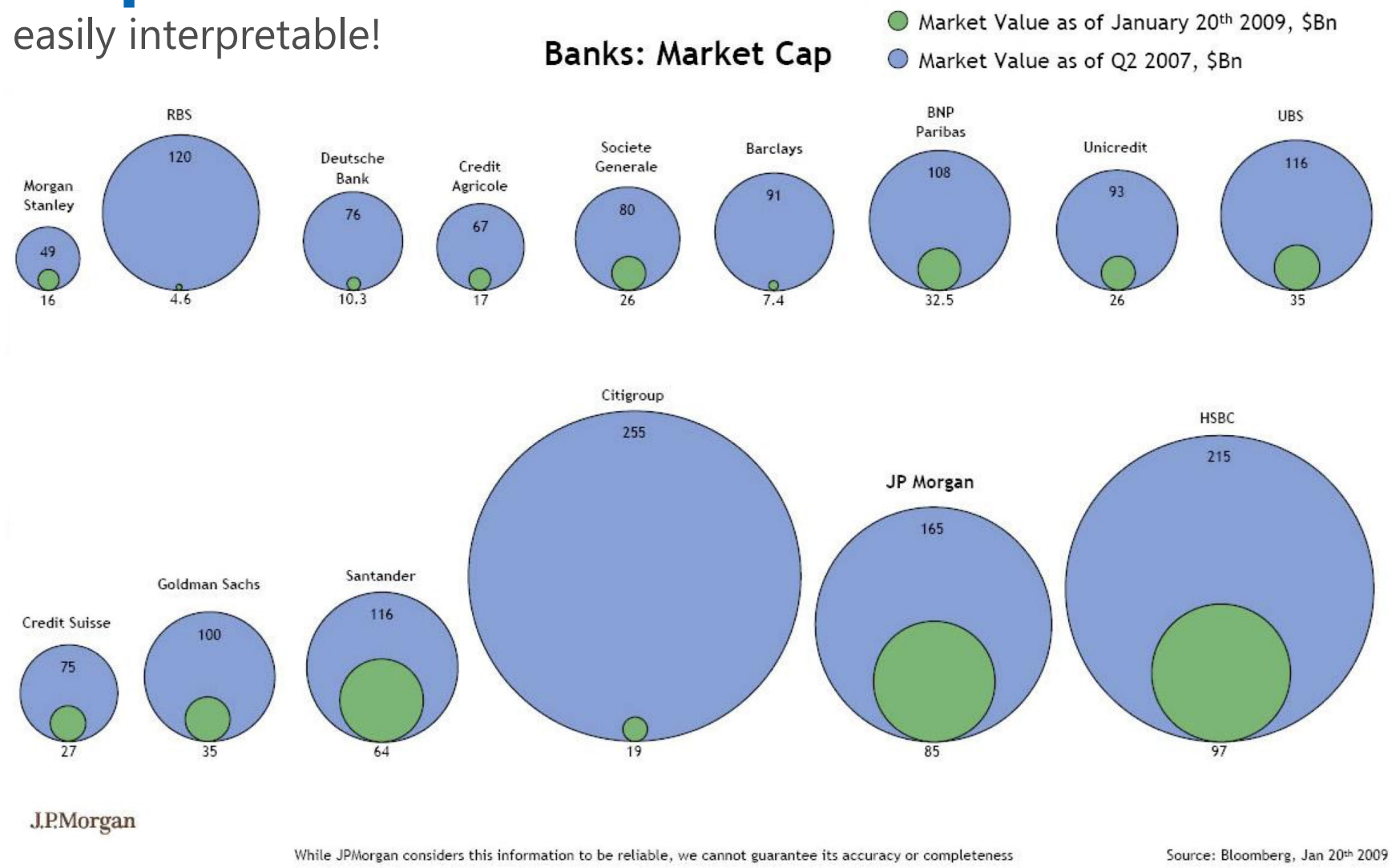


Source: Giga Research, a wholly owned subsidiary of Forrester Research, Inc.

Source: 2018, Stephen Few, PerceptualEdge

Not quite a pie chart

2D-Area is not easily interpretable!

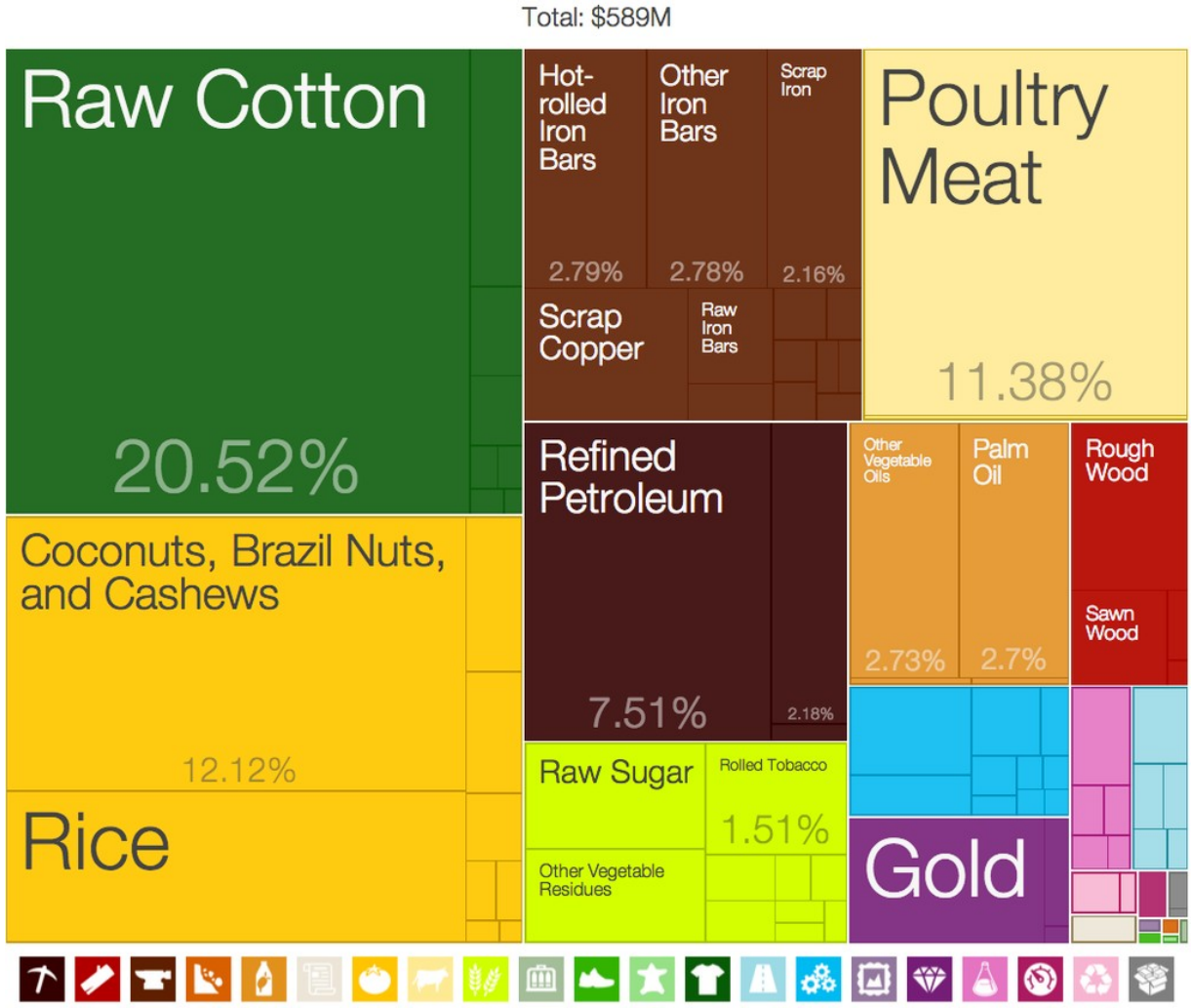


Source: 2018, Stephen Few, PerceptualEdge

Tree maps – (Imperfect) Alternative to Pie Chart

Notice what happens when we have many sub-categories

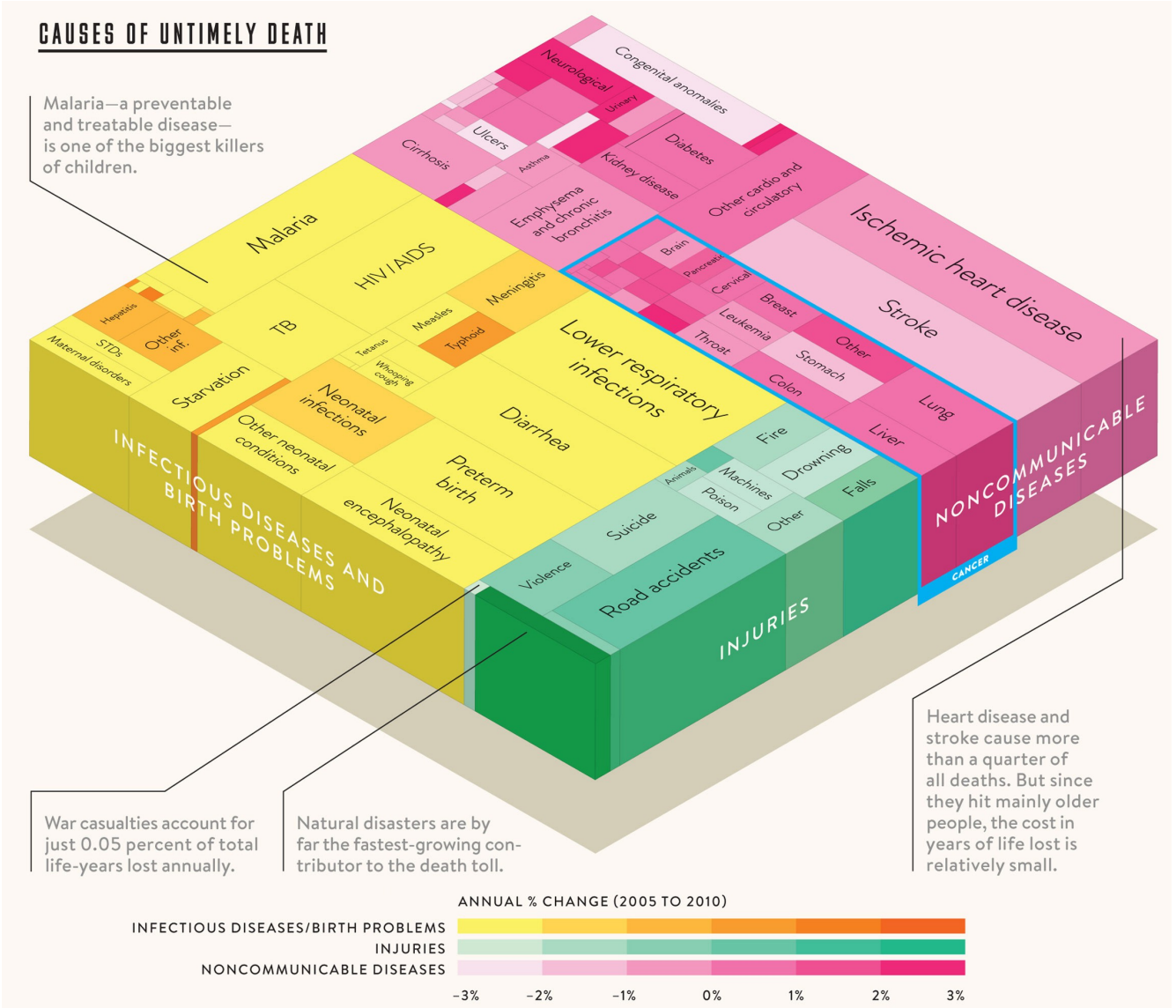
- Tree maps **display hierarchical data** as a set of nested rectangles.
- Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing sub-branches.
- A leaf node's rectangle has an area **proportional to a specified**.
- Often the leaf nodes are coloured to show a **separate dimension** of the data
- Tree maps use space **efficiently compared to pie charts,** but we cannot legibly read **small sub-categories**



Causes of Untimely Death

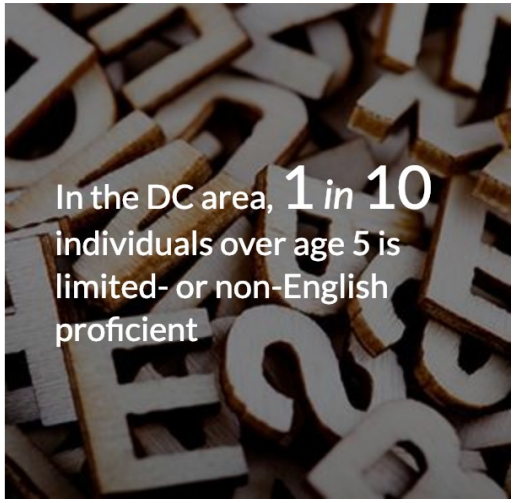
Bill Gate's favorite chart

Notice the color scale at the bottom



Text – Impact Metrics

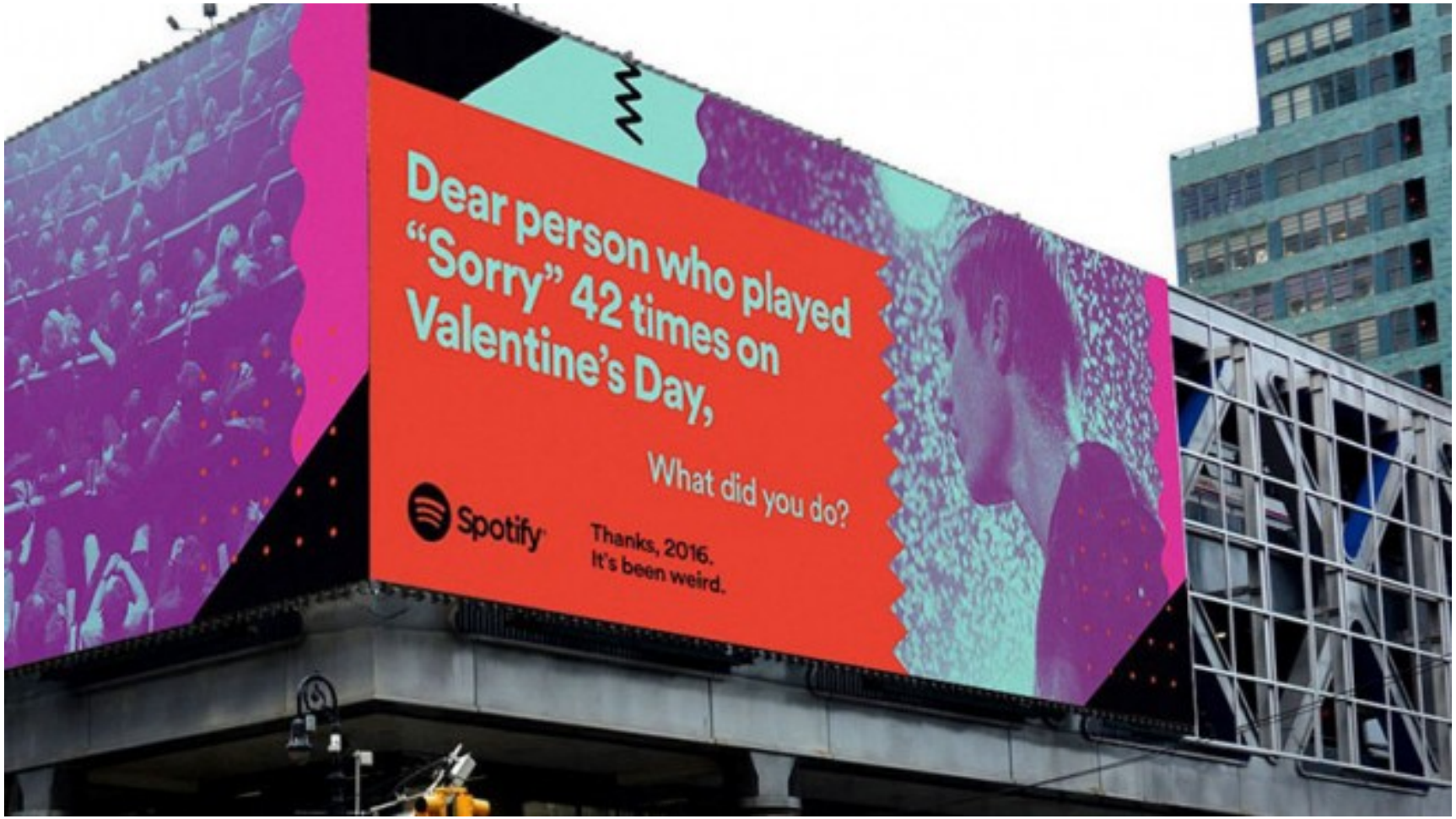
Draw attention with large font, place a semi-opaque background image for context.



Source: https://www.urban.org/data-viz

Text – Text ads in the wild

Lessons from Spotify
– less is more



Choosing a Visual - Table

Never use this in presentations

Product	Category	Profit Margin	Average Price
A	Furniture	15%	123.1
B	Bedding	20%	198.1
C	Office Supplies	10%	212.2
D	Home Decor	25%	432.5

Choosing a Visual - Table

Never use this in presentations

1 Remove/lighten **borders**

Product	Category	Profit Margin	Average Price
A	Furniture	15%	123.1
B	Bedding	20%	198.1
C	Office Supplies	10%	212.2
D	Home Decor	25%	432.5

2 **Left** align **text**.

3 **Right** align **numbers**.

Choosing a Visual - Heatmap

Use both sides of the brain for impact

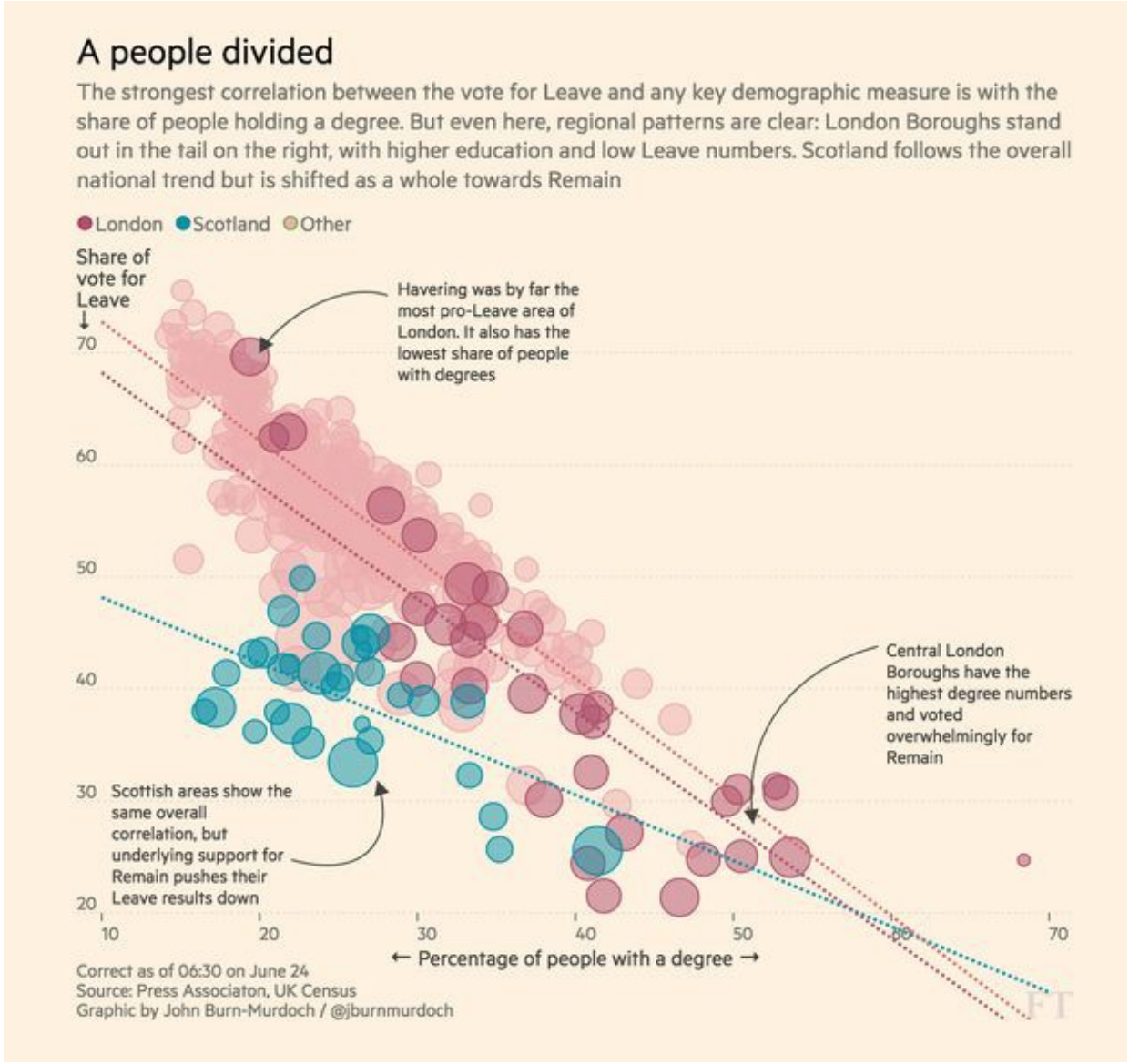
Product	Category	Profit Margin	Average Price
A	Furniture	15%	123.1
B	Bedding	20%	198.1
C	Office Supplies	10%	212.2
D	Home Decor	25%	432.5

4 Color saturation

Scatterplots
Is having a degree correlated to voting ‘Leave’

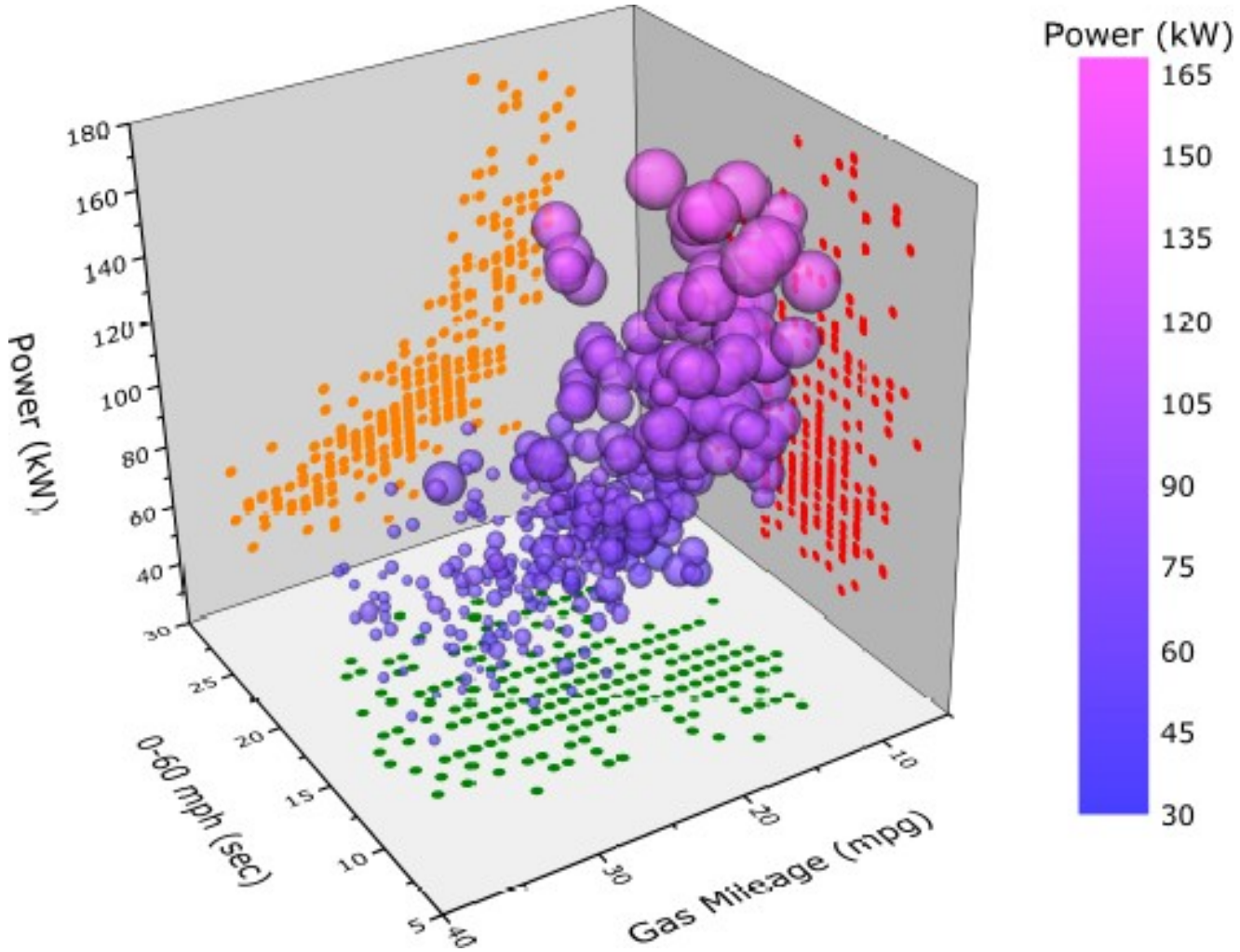
Best Practice

Pre-attentive attributes
Two dimensions maximum
Use color to differentiate



Scatterplots
How many dimensions do we have here?

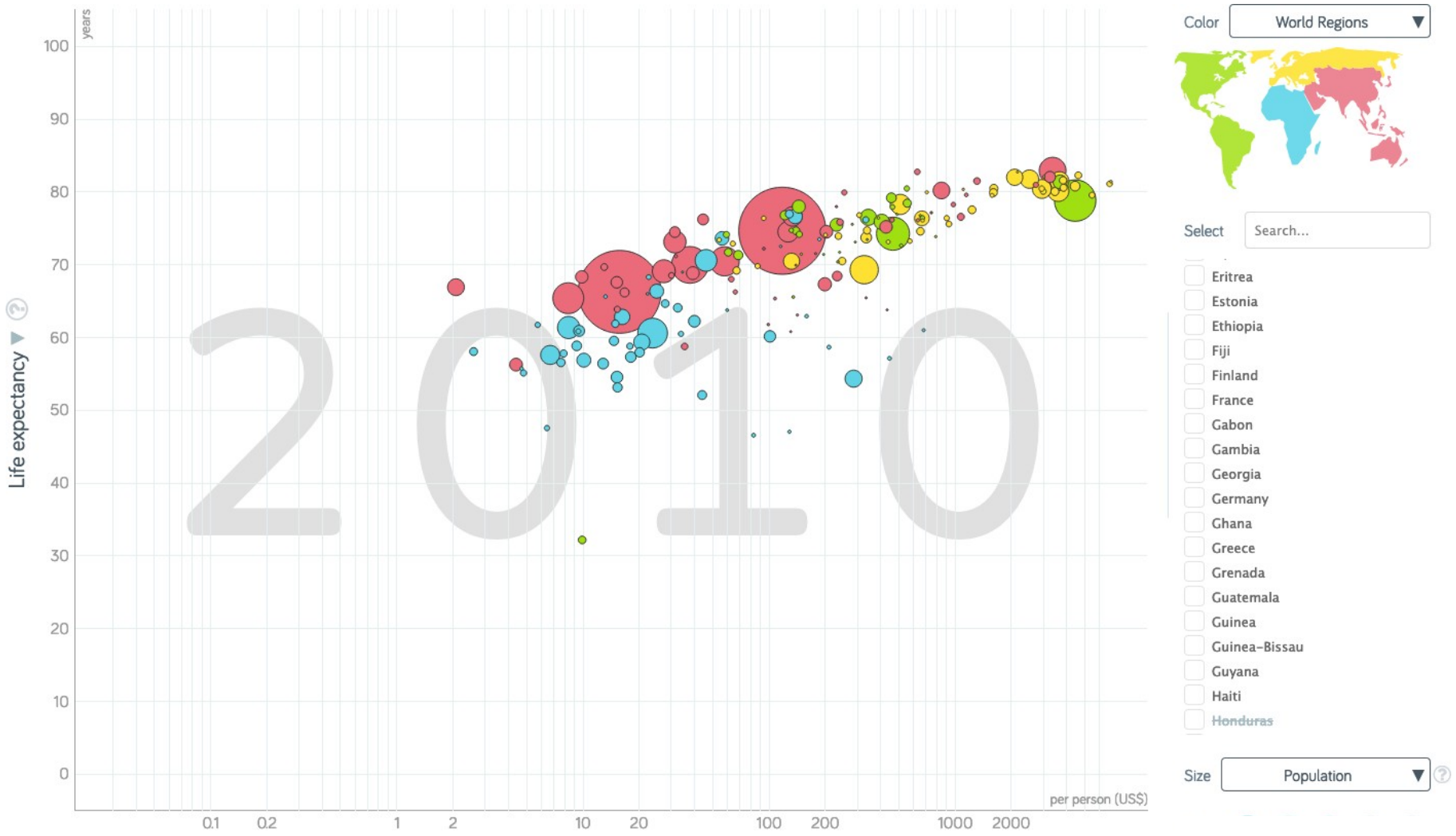
Features about a car plotted in a multi-dimensional scatter plot.



Reverse causation:

Countries which spend more on health care tend to have longer life expectancy

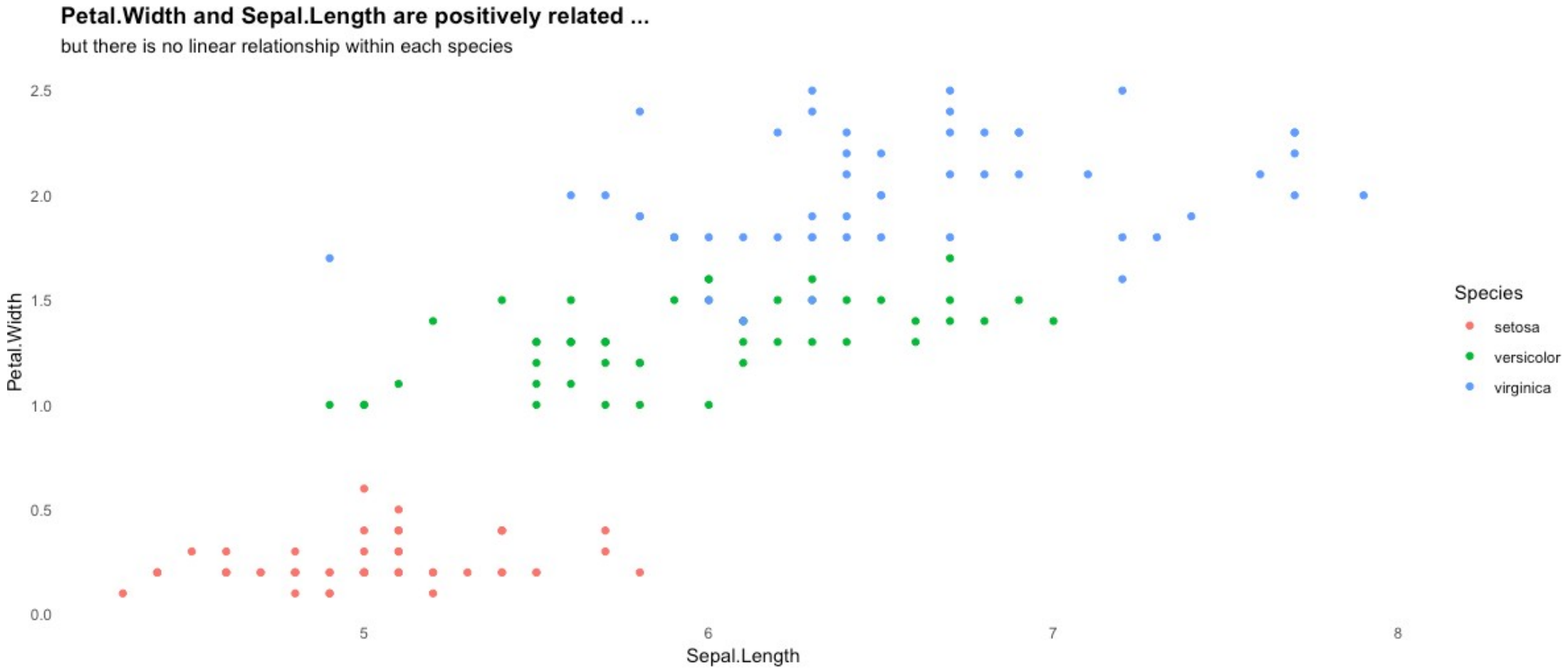
But what if countries spend more on health care because they have lower life expectancies? (Filmer & Pritchett 1999)



Gapminder

Simpson's Paradox in the Iris Dataset

Recall Day 2 of Excel Analytics



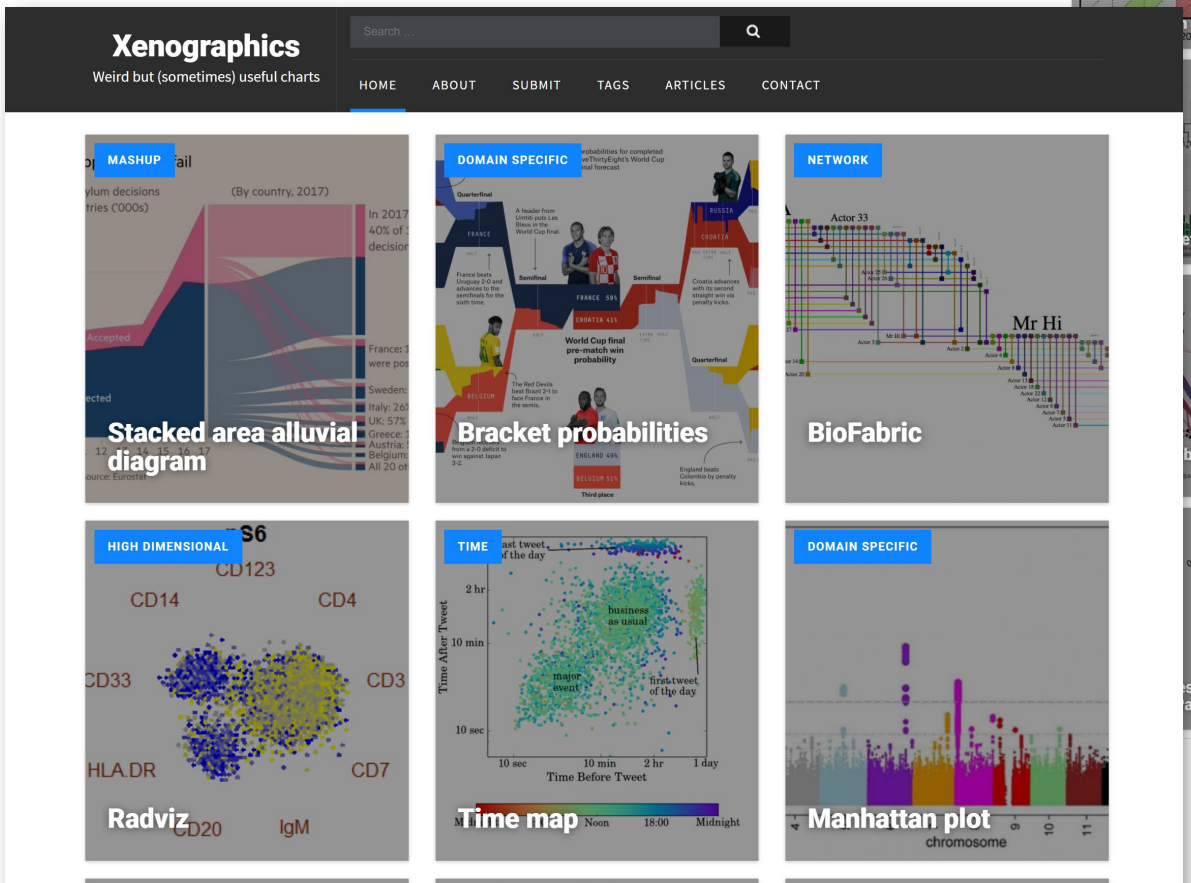
Choosing a Visual

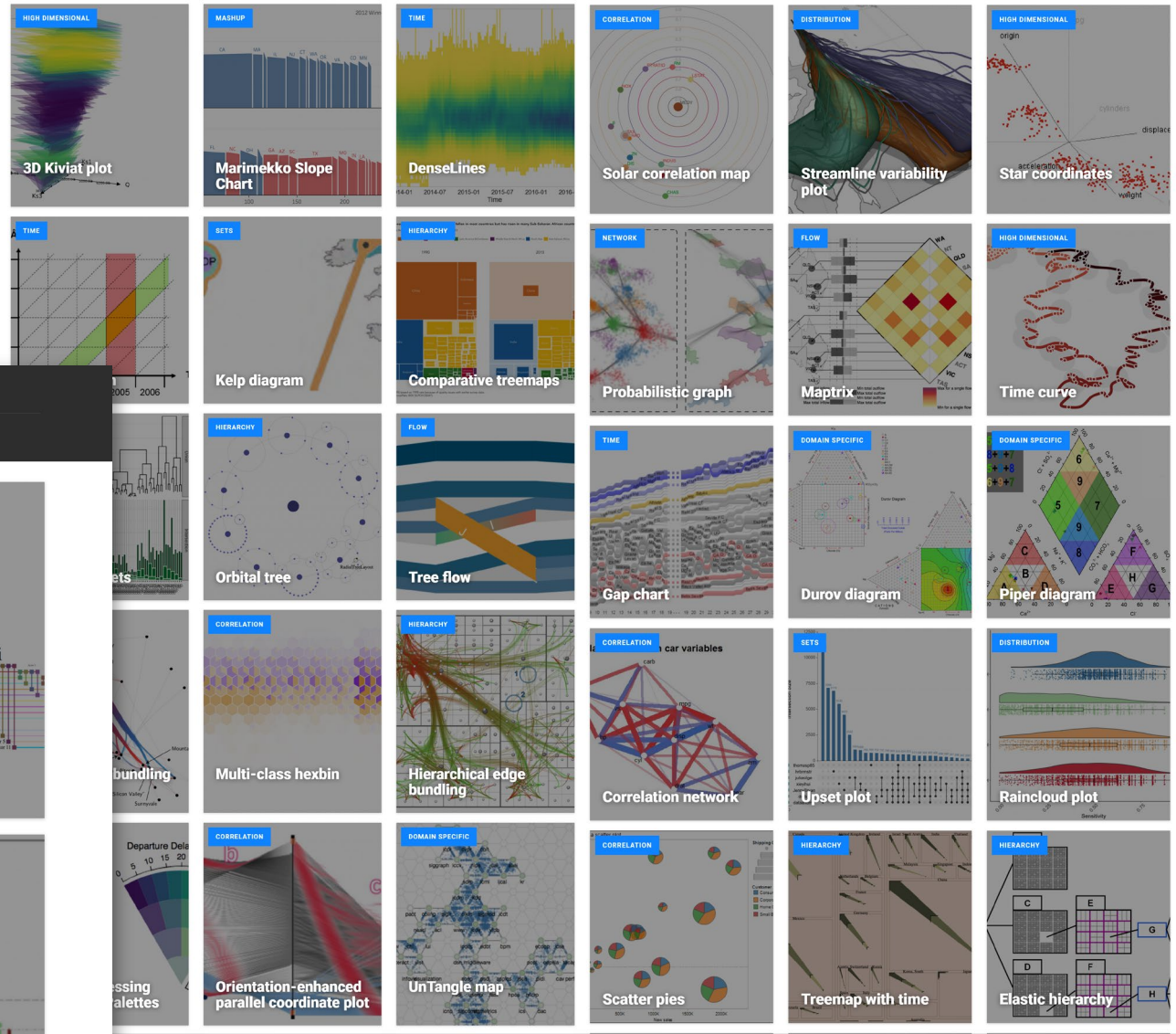
Form follows function.

Choose a visual that supports the ‘What’

Weird Charts' can be great but..

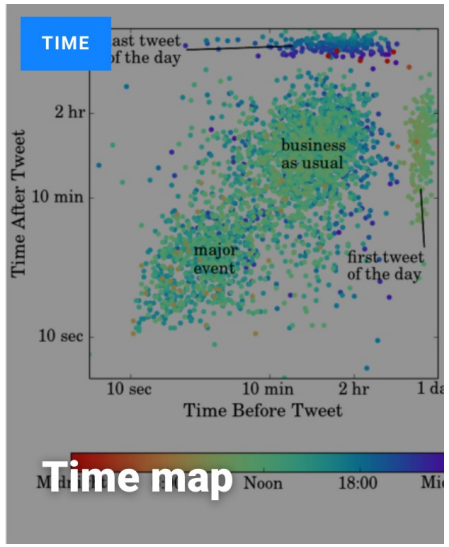
Master the basics first!
https://xeno.graphics/

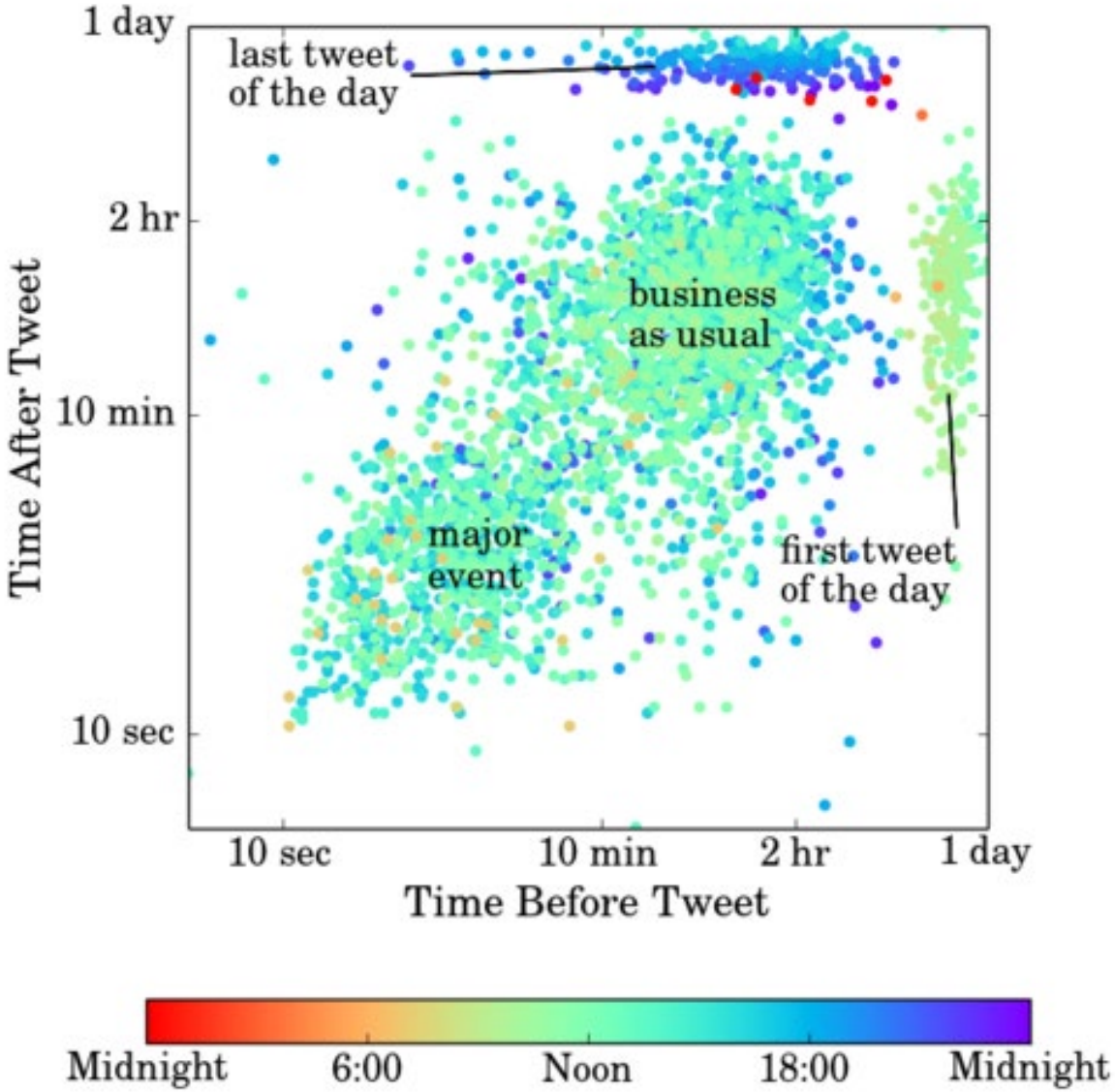




Time Map
Not Recommended
Why?

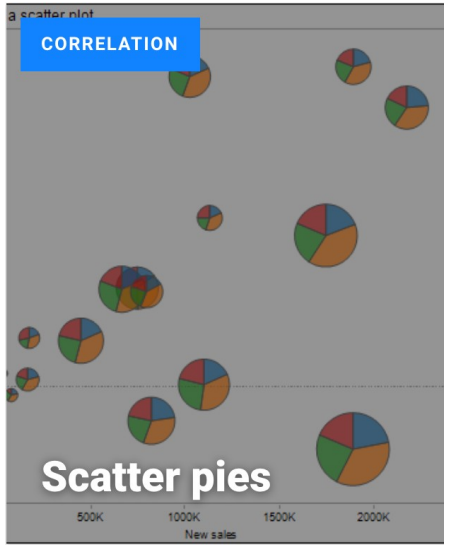
Should only use as many dimensions in your visual as you have in your data.

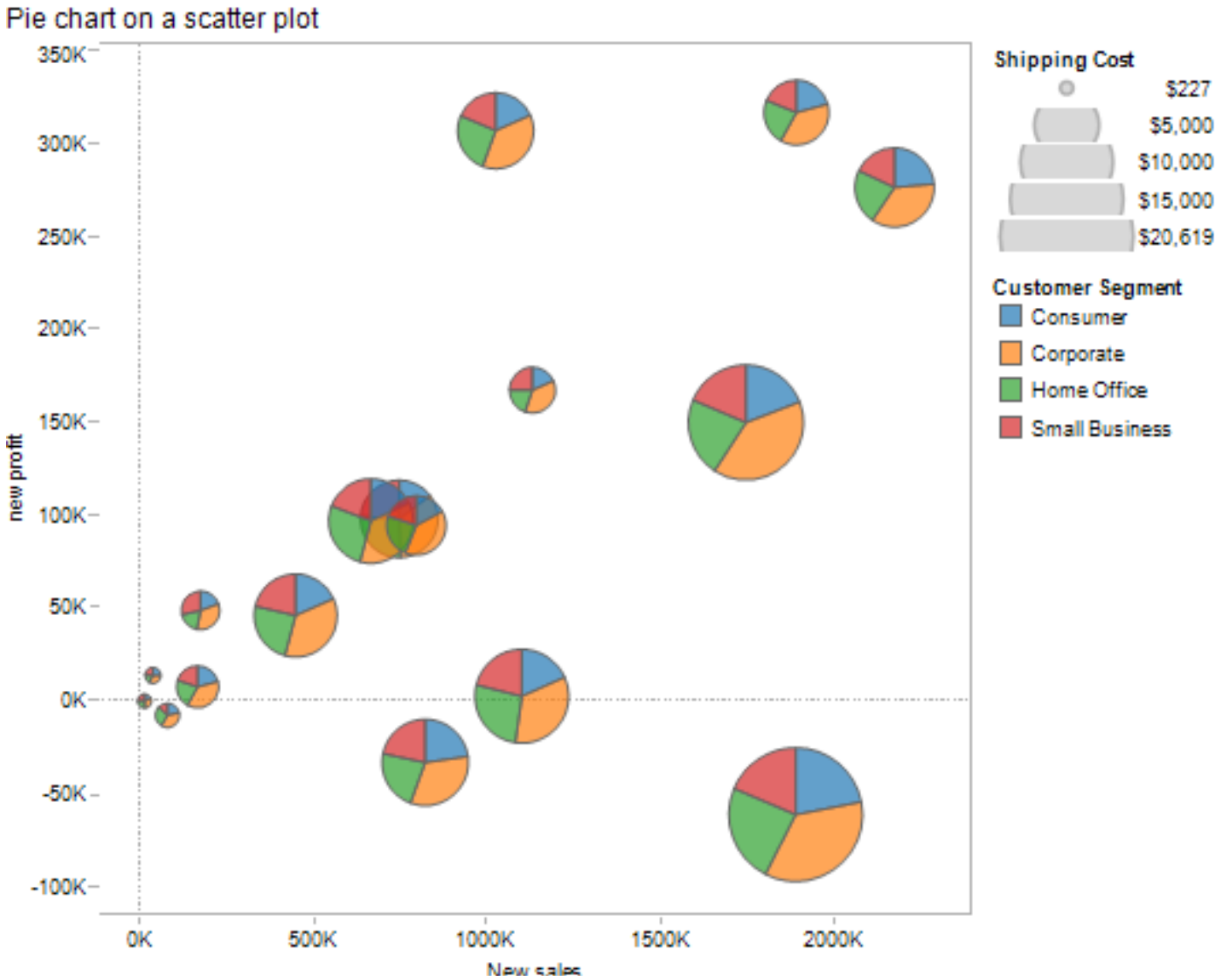




Scatter Pies
Not Recommended
Why?

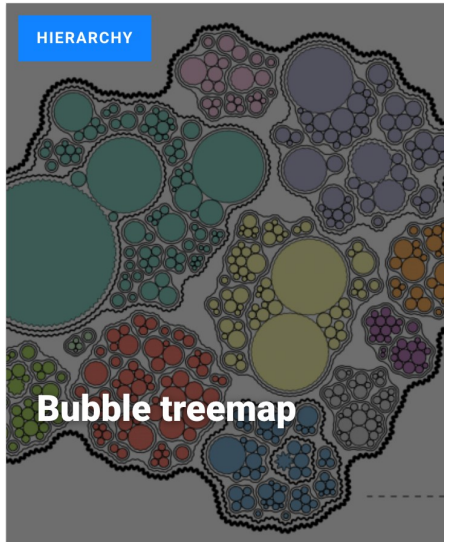
"The only thing worse than a pie chart is several of them" – Edward Tufte

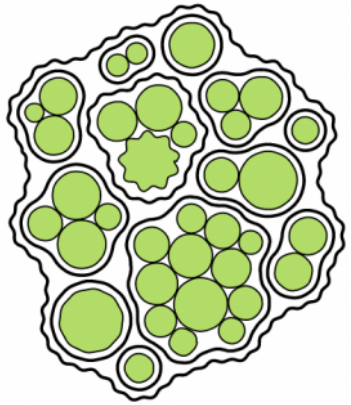


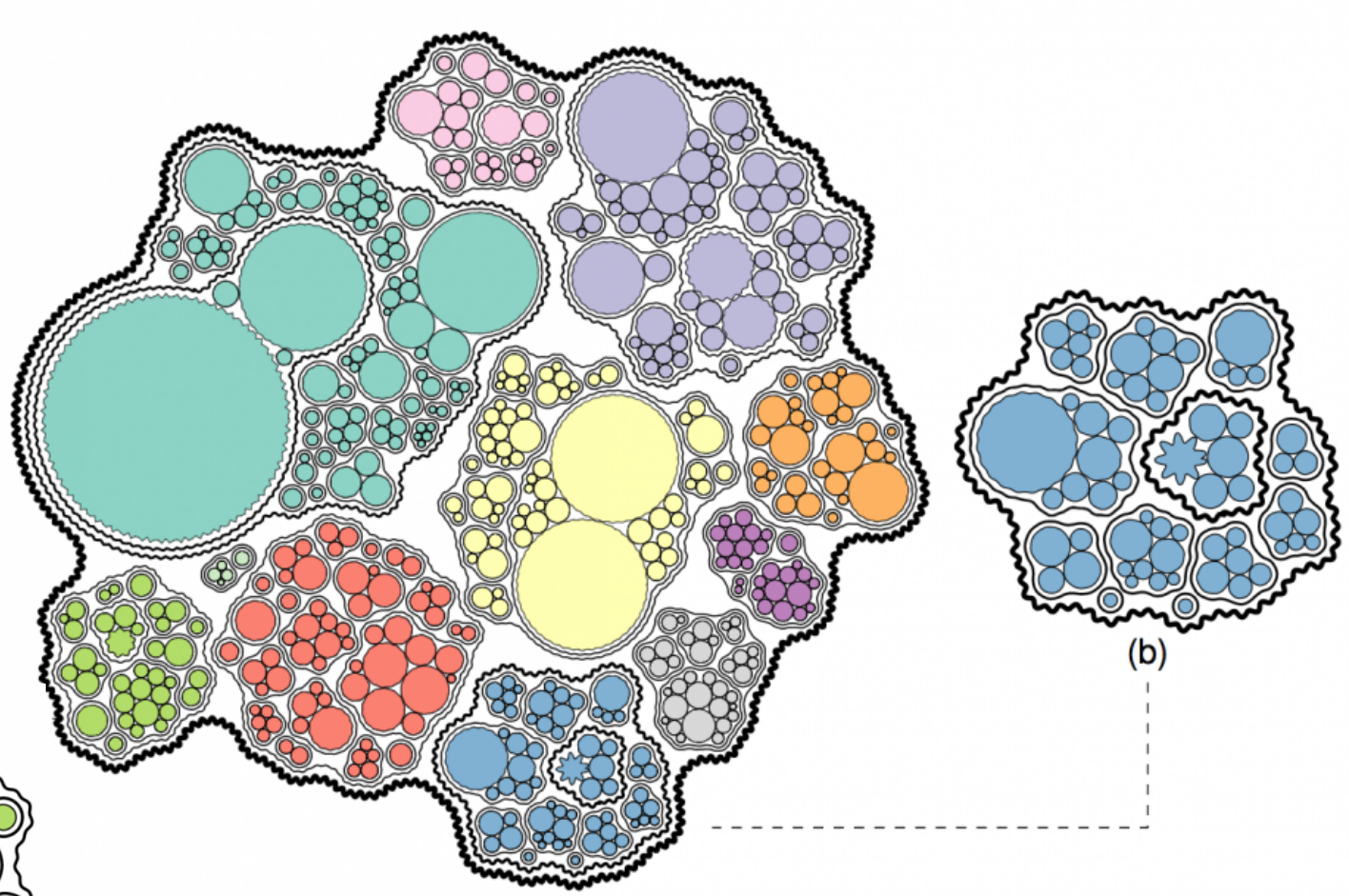


Bubble treemap
Not Recommended
Why?

How large is each bubble?
There is no space to label these bubbles.







Gestalt Principles of Visual Grouping

Borrowing from Cognitive Science

- How individuals **perceive order** in the world around them
- Identifying which elements in our visuals are signal and which might be noise
- Six principles:
  - Proximity
  - Similarity
  - Closure
  - Continuity
  - Connection
  - Enclosure

Gestalt Principles of Visual Grouping

Proximity

Objects that are physically close together are perceived as belonging together

We can influence the perceived orientation of a table simply by changing the spacing between some dots

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

Gestalt Principles of Visual Grouping

Similarity

Objects that are of similar color, shape, size, or orientation are perceived as related or belong to part of a group

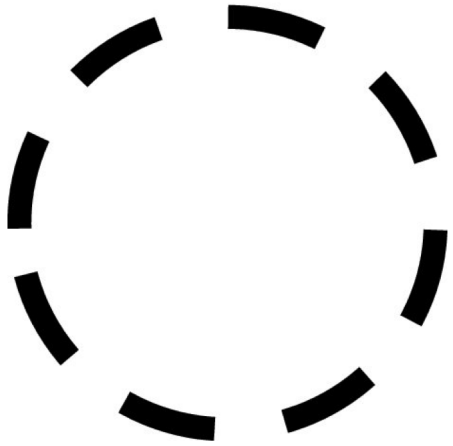
We can use color to guide a reader's eye while scanning a table

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

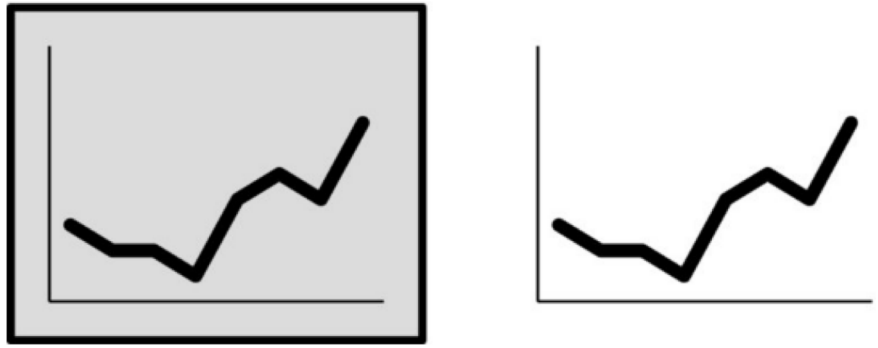
Gestalt Principles of Visual Grouping

Closure

We like things to be simple and fit in the constructs that are already in our heads. We tend to perceive a set of individual elements as a single, recognizable shape when we can.



We can remove chart borders and background shading and our graph will still appear as a cohesive entity

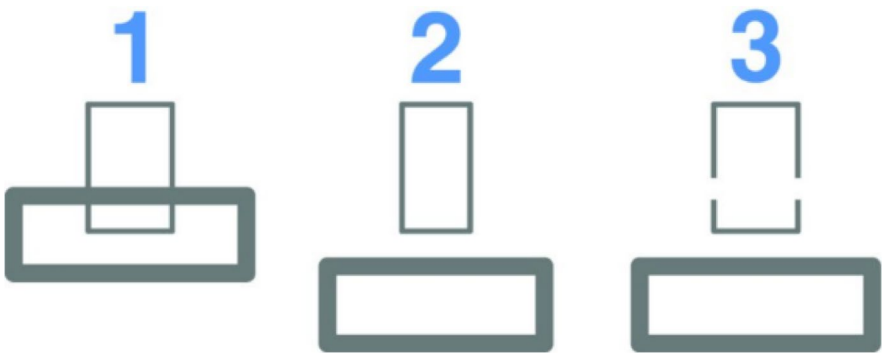


Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

Gestalt Principles of Visual Grouping

Continuity

Similar to closure: When looking at objects, our eyes seek the smoothest path and naturally create continuity in what we see even if it may not explicitly exist



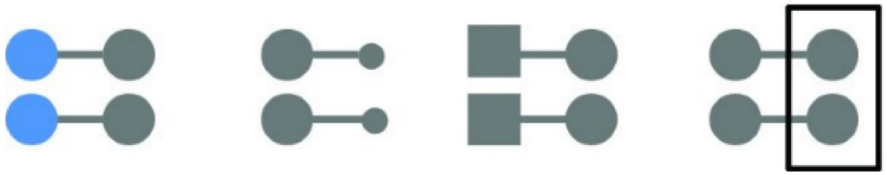
Here we have removed the y-axis and we can still see the bars aligned because of consistent white space between the labels and the data

Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

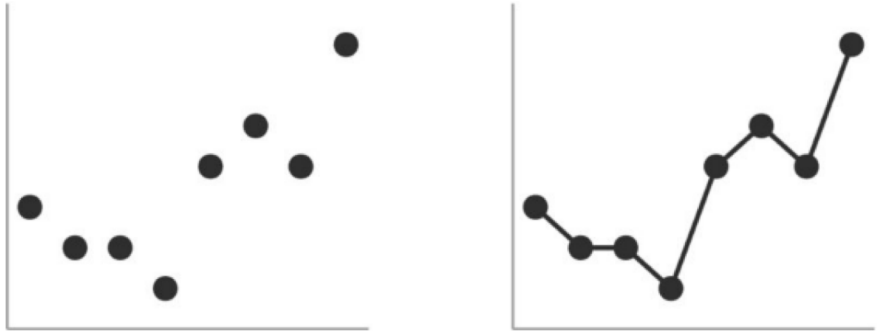
Gestalt Principles of Visual Grouping

Connection

We tend to think of objects that are physically connected as part of a group. The connectivity property typically has a **stronger** associative value than similar color, size, or shape.



We use this principle in line graphs to help our eyes see order in the data.

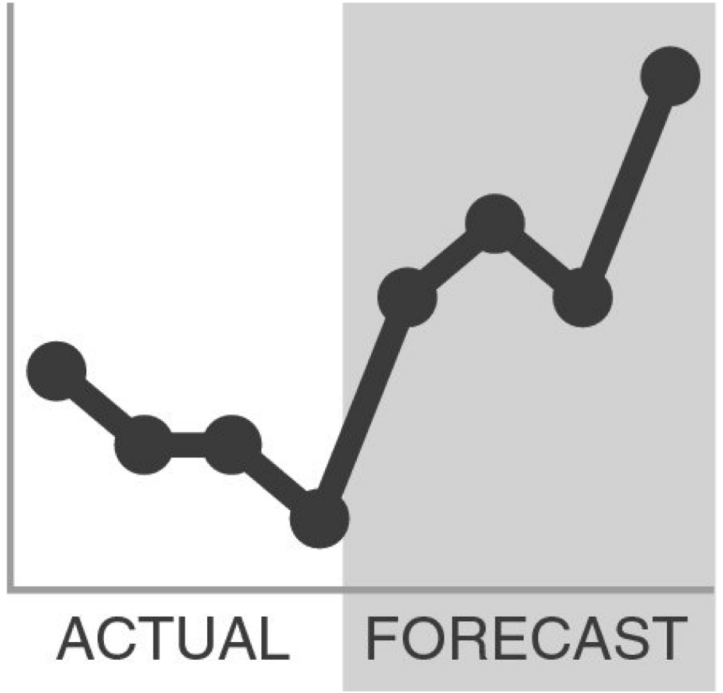


Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

Gestalt Principles of Visual Grouping

Enclosure

We think of objects that are physically enclosed together as belonging to part of a group

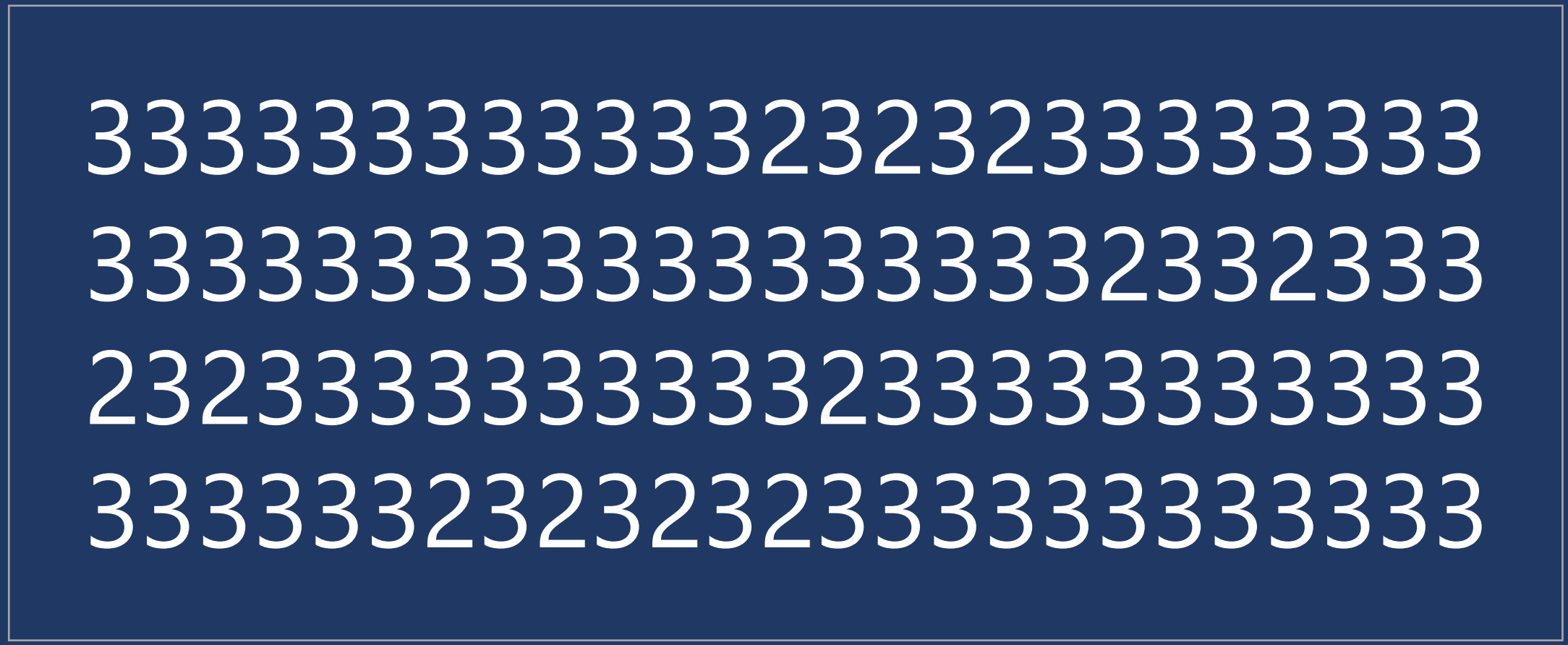


Source: Harvard CS109 – Hanspeter Pfister and Joe Blitzstein http://cs109.org

Visual Hierarchy

Pre-attentive attributes

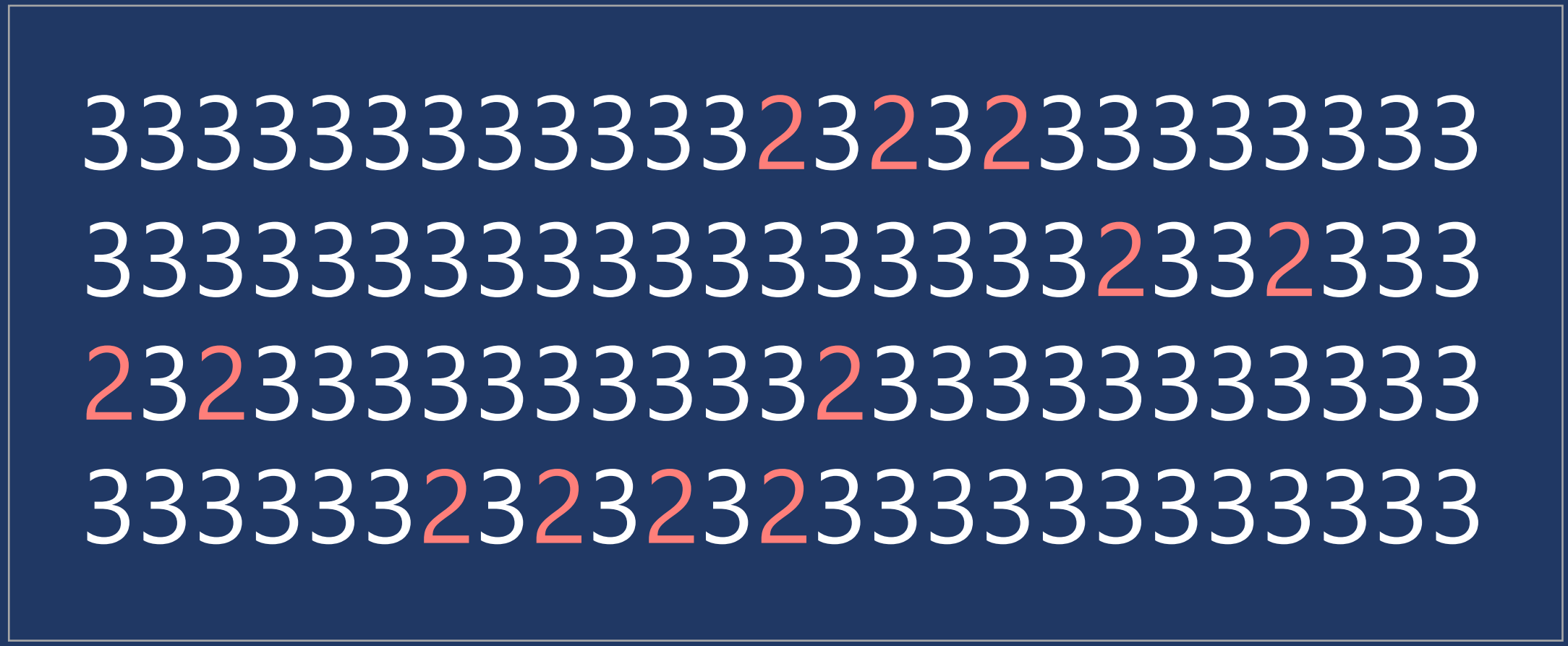
How many 2's?



Visual Hierarchy

Leveraging Pre-attentive attributes

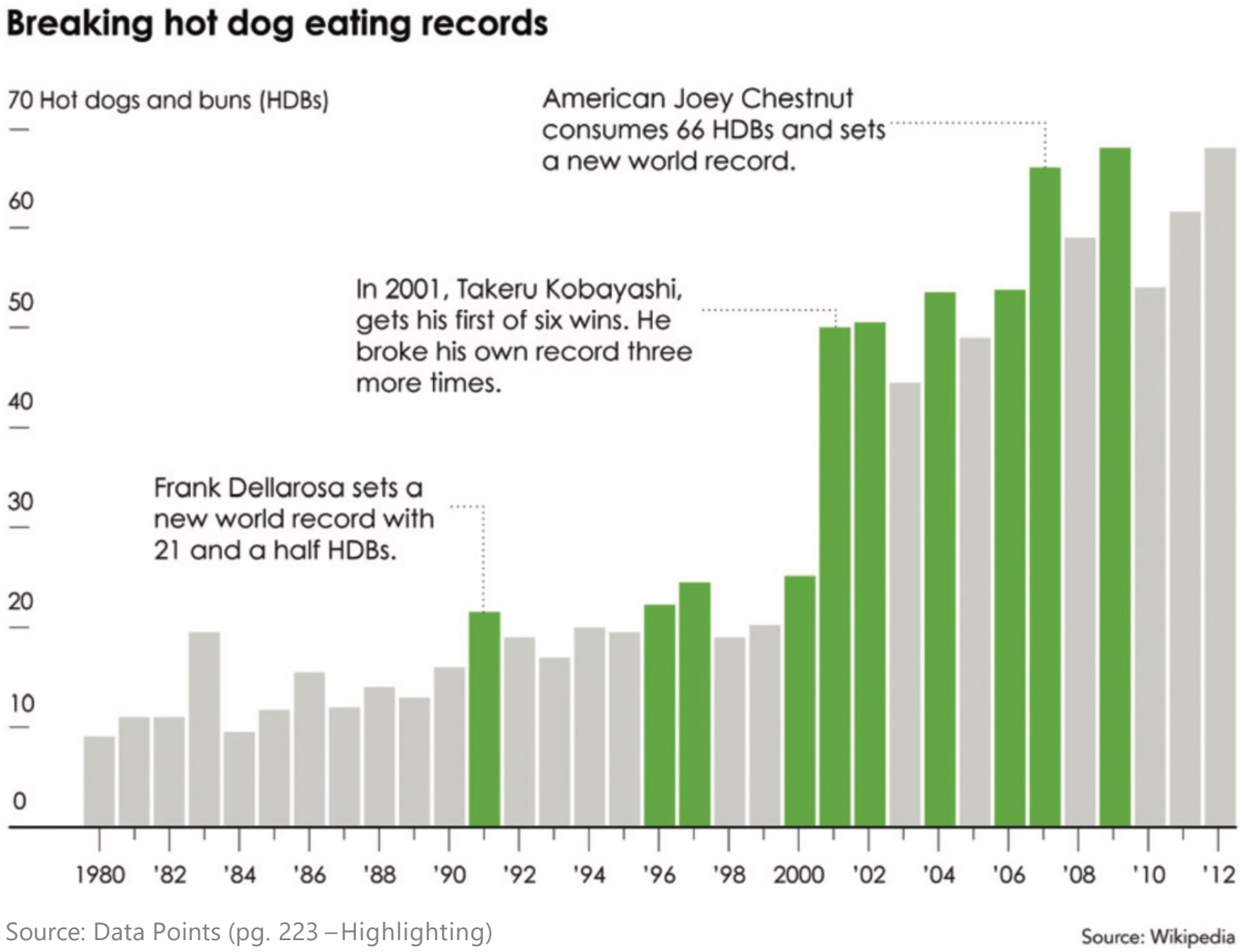
How many 2's?



Visual Hierarchy

Think about how your audience will process.

- Bold title
- Smaller font subtitle
- Green bar highlight
- Axes ticks small font
- Annotations



Source: Data Points (pg. 223 – Highlighting)

Anatomy of Text Hierarchy

Adding order to chaos.

Subtitle supports the main title.

Title text is larger and the most dominant text on the page.

Web links can be color-coded, underlined, or both.

Body copy is a neutral color and smaller than headline copy, so headlines and subheads stand out.

Colored box is great for emphasized text or pull quotes.

Footers are the least dominant element on the page.

Tabulated navigation helps identify where you are; colored text identifies the current page.

Source: Slidedocs – Nancy Duarte (2014)

Chart Junk – Wasted space / graphics on a visual

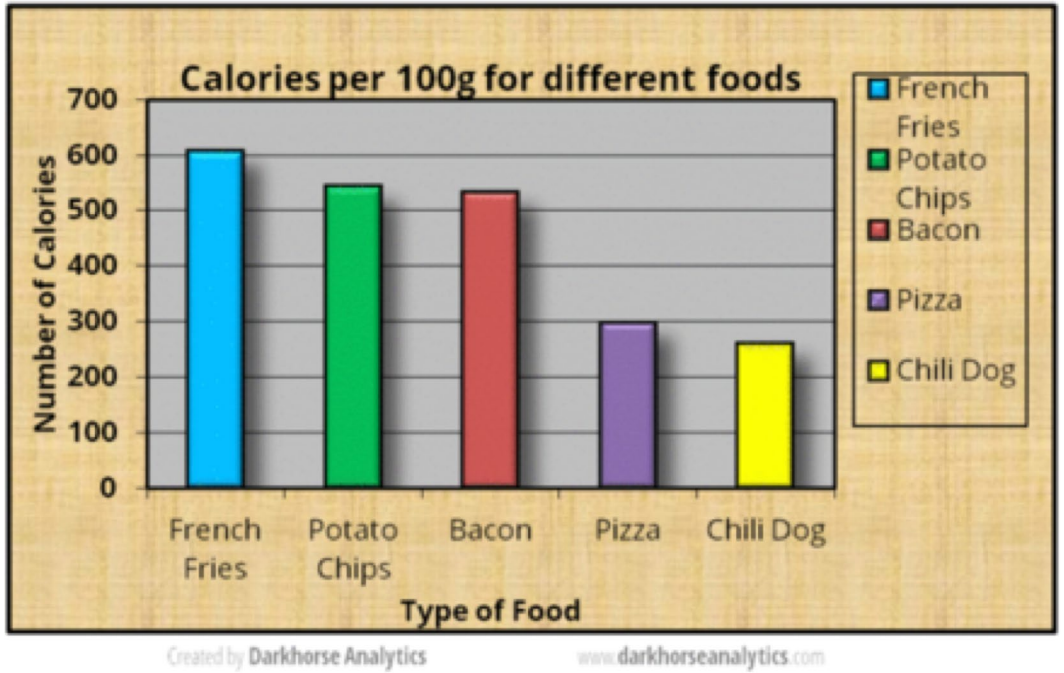
Note: You won't worry about this if you have a clear story

"Data is **boring**, so I need to make an attractive visual to grab the audience's attention"

Guided Redesign

Less is more!

Remove backgrounds



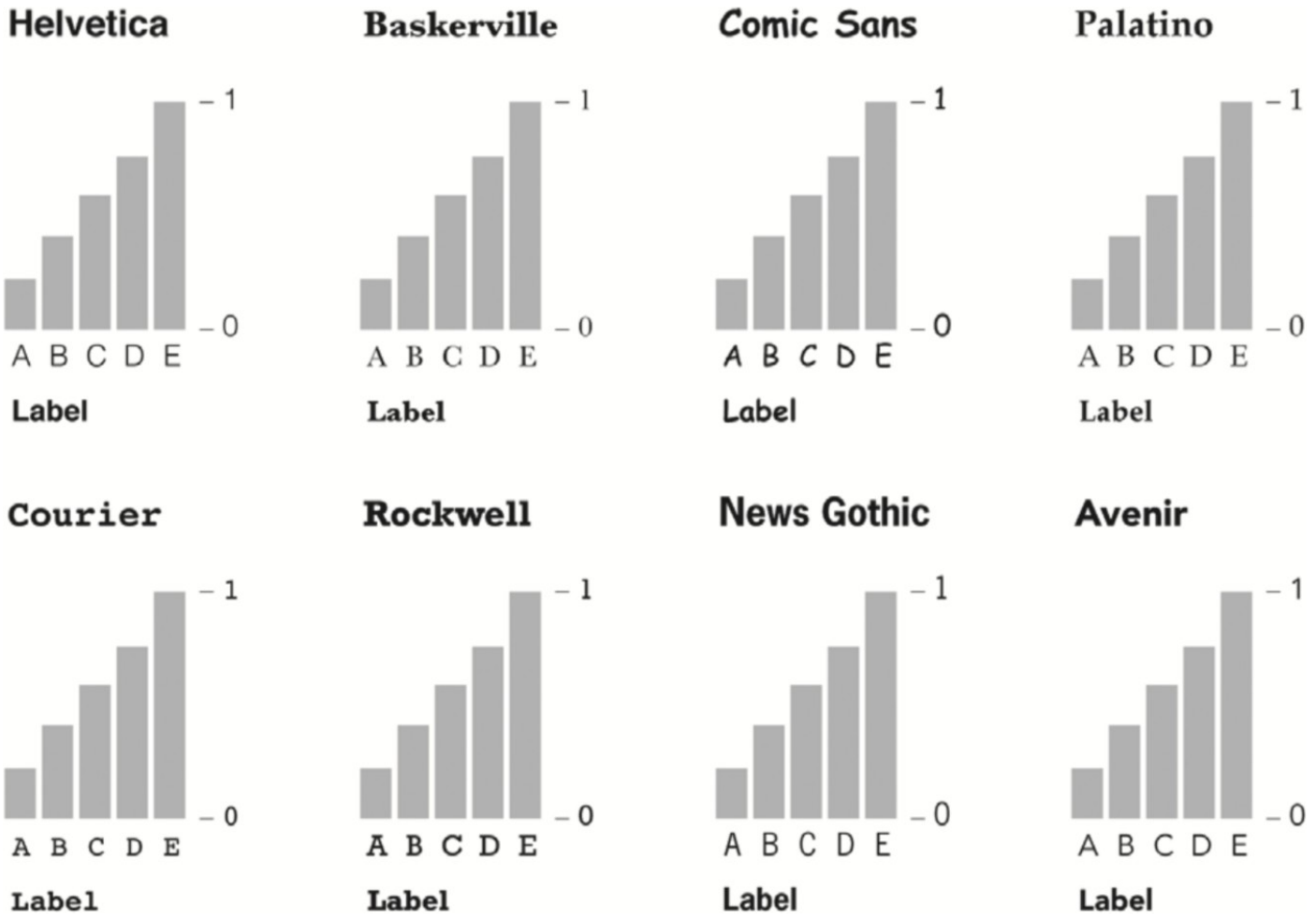
Design Thinking

Edward Tufte prefers Serif – what do you prefer

Sans-serif	Arial	Serif	Times New Roman
Caps lock	ARIAL		TIMES NEW ROMAN

Source: The Visual Display of Quantitative Information (pg. 187) – Edward Tufte

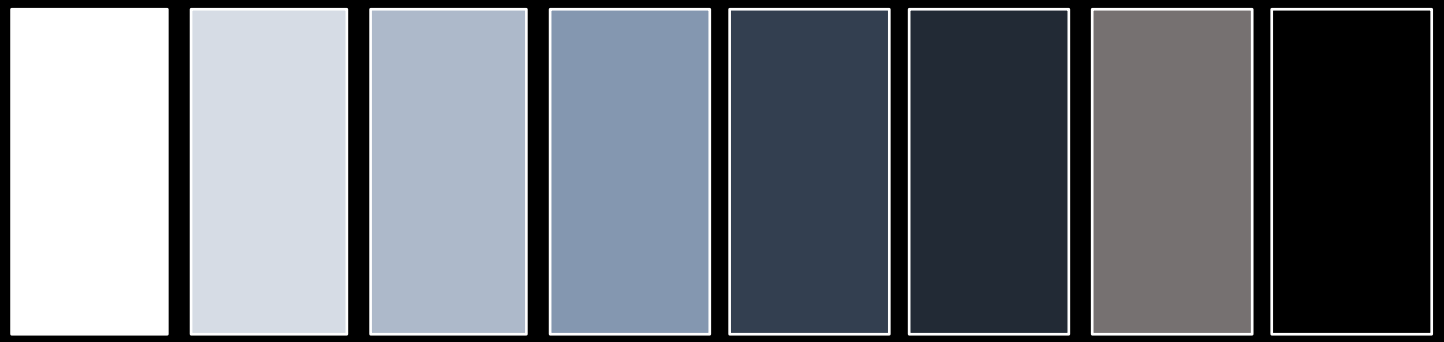
Design Thinking
Fonts & Feels



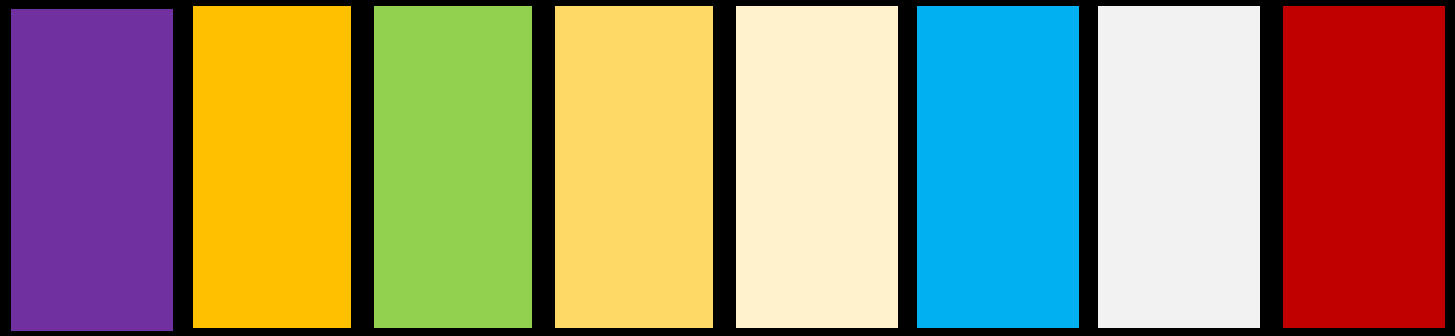
Source: Data Points (pg. 234 – Experiment with Typography)

Design Thinking

Do not use more than 5-8 colors at once.

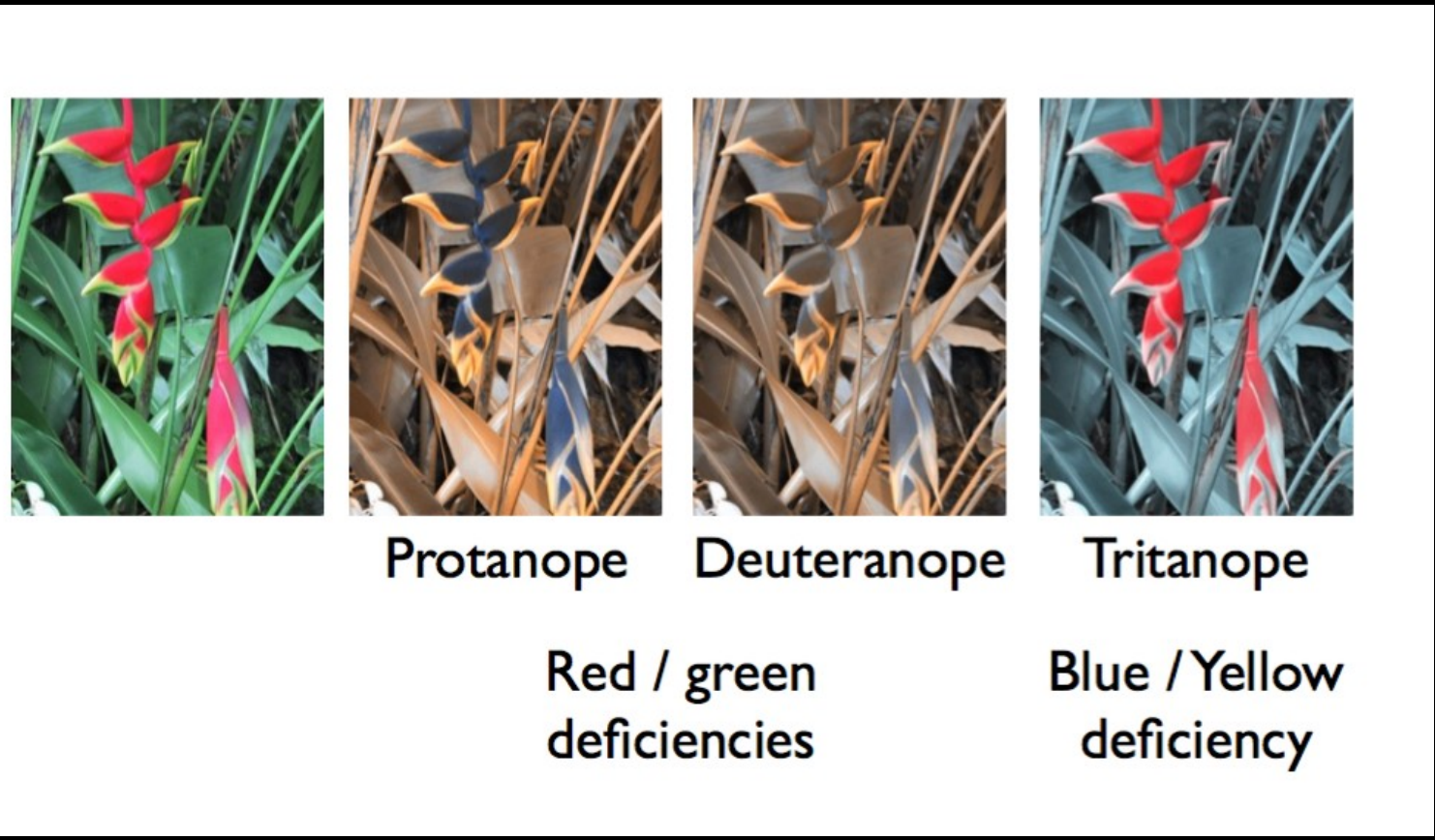


Use Color Purposefully



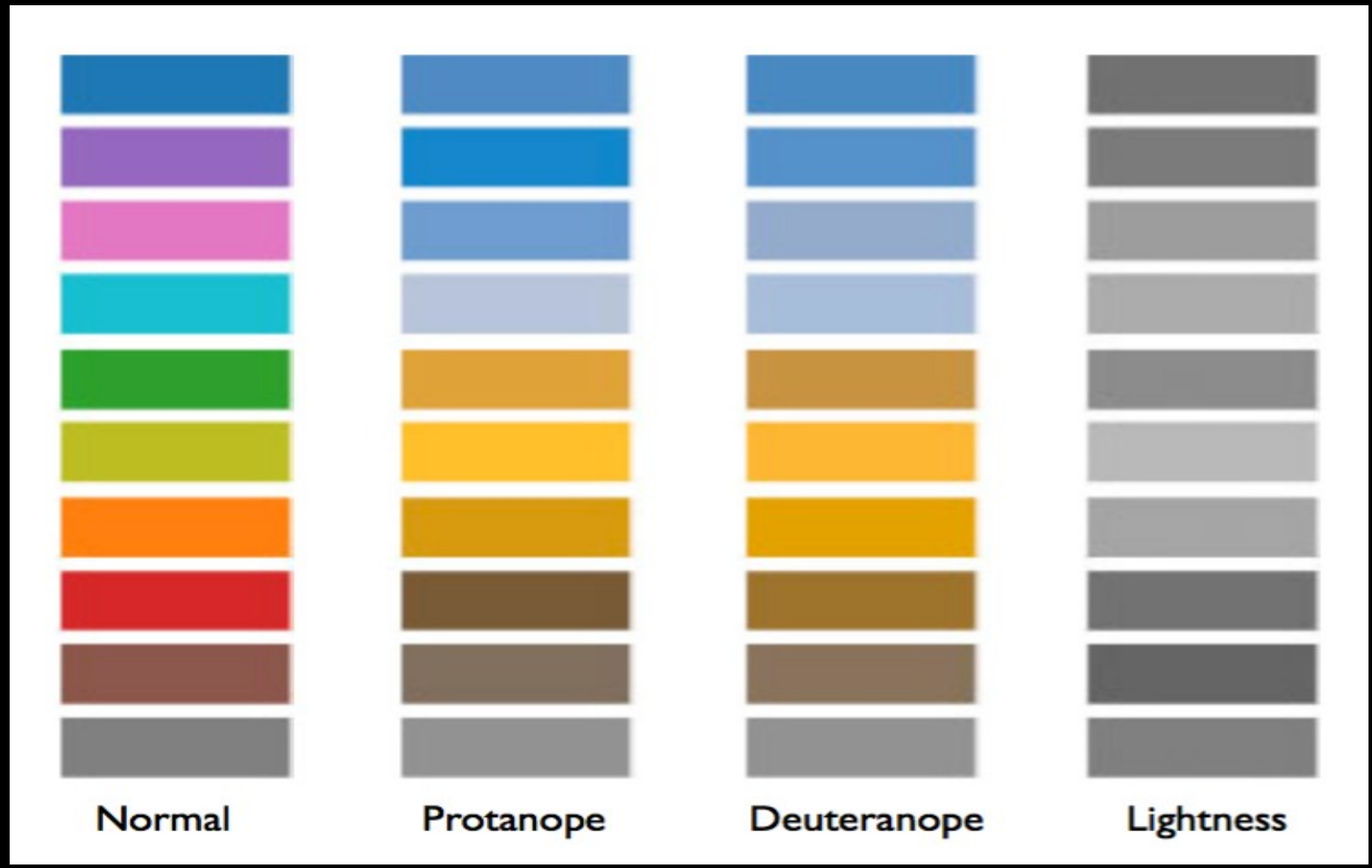
Design Thinking

Color Blindness



Design Thinking

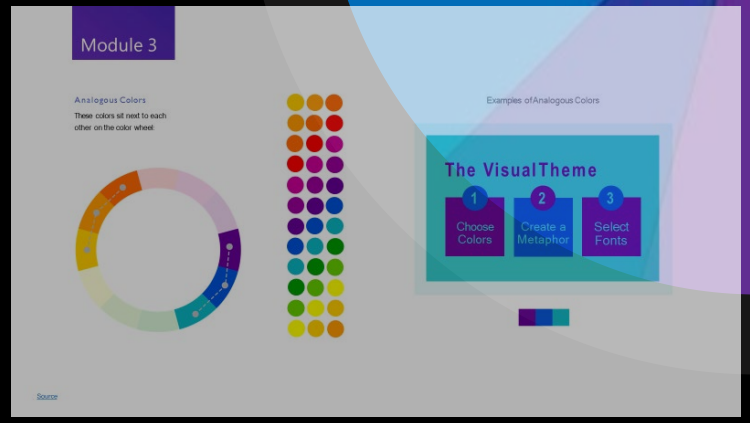
Color Blindness



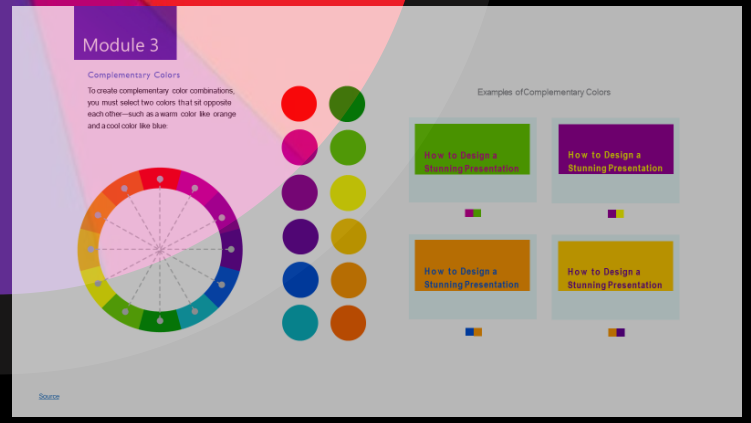
Design Thinking

Use the Color Wheel

Monocratic Color



Analogous Color



Complementary Color

Design Thinking

Opposites attract

1. Opposite colors on the color wheel.

2. These are also less-susceptible to color-blind interpretation.

Leverage White Space
Reduce cognitive overload and improve focus.

The more breathing room you give to a design or product, the more luxurious it's perceived to be.

Examples of advertisements using white space

Source: Slidedocs – Nancy Duarte (2014)

Convince the headmaster the summer science camp was a success

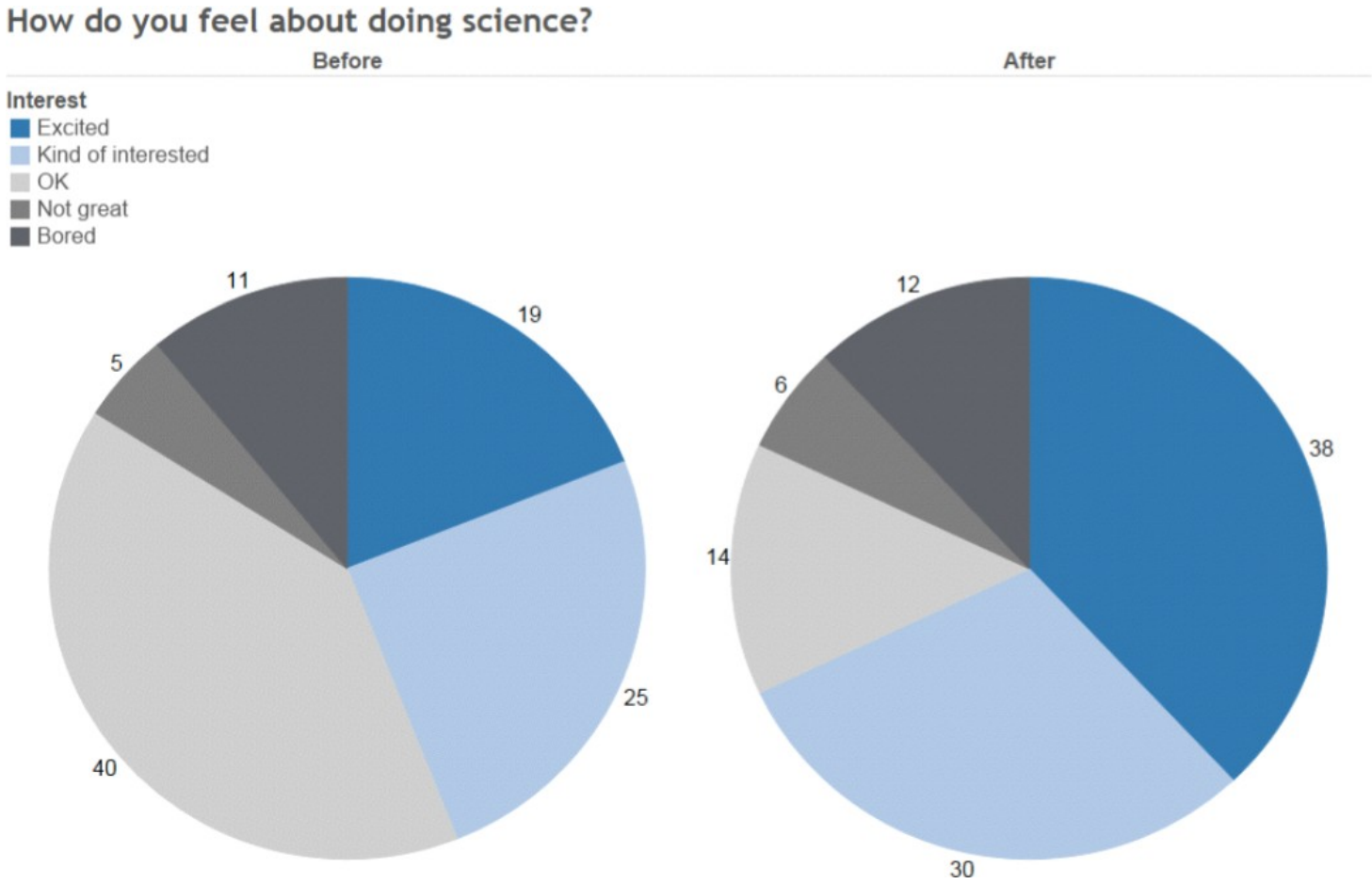
Think about which chart will work best for the data we have here

Table

Interest	Before	After
Excited	19	38
Kind of interested	25	30
OK	40	14
Not great	5	6
Bored	11	12

Source: Storytelling with Data – Cole Nussbaumer Knafflic

Design Activity
Sample visual 1
Pie Chart

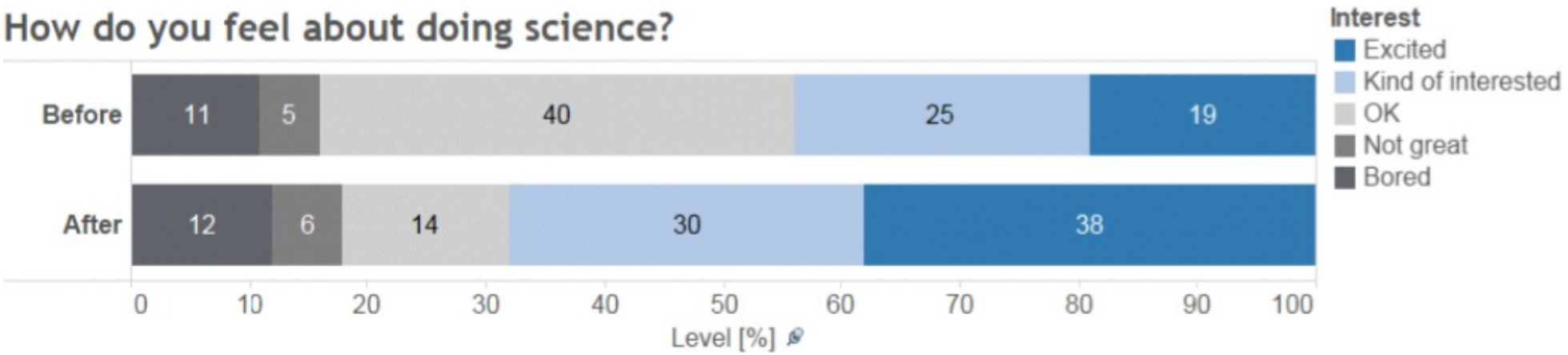


Source: Storytelling with Data – Cole Nussbaumer Knafflic

Design Activity

Sample visual 2

100% Stacked Horizontal Bar Chart

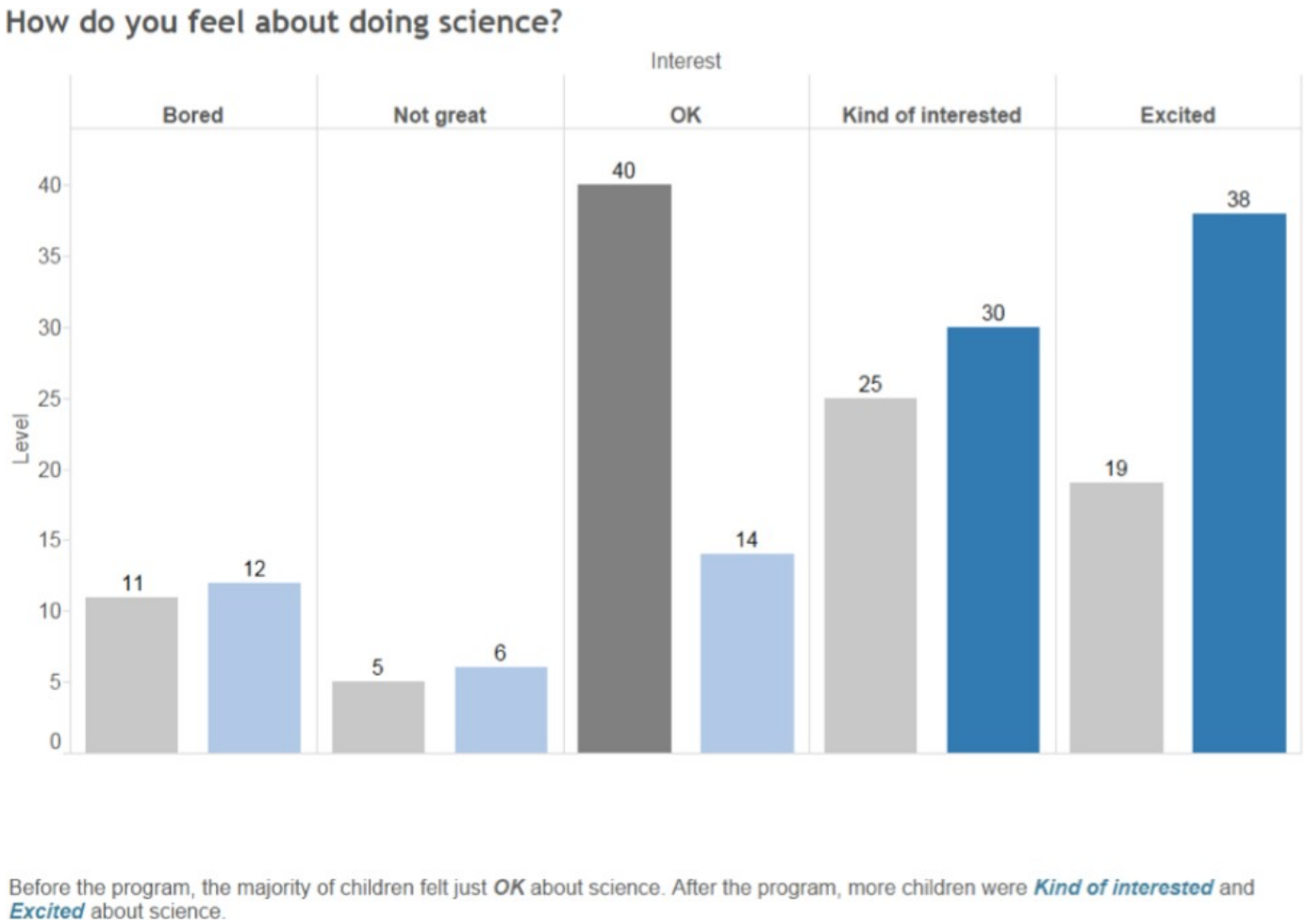


Source: Storytelling with Data – Cole Nussbaumer Knafflic

Design Activity

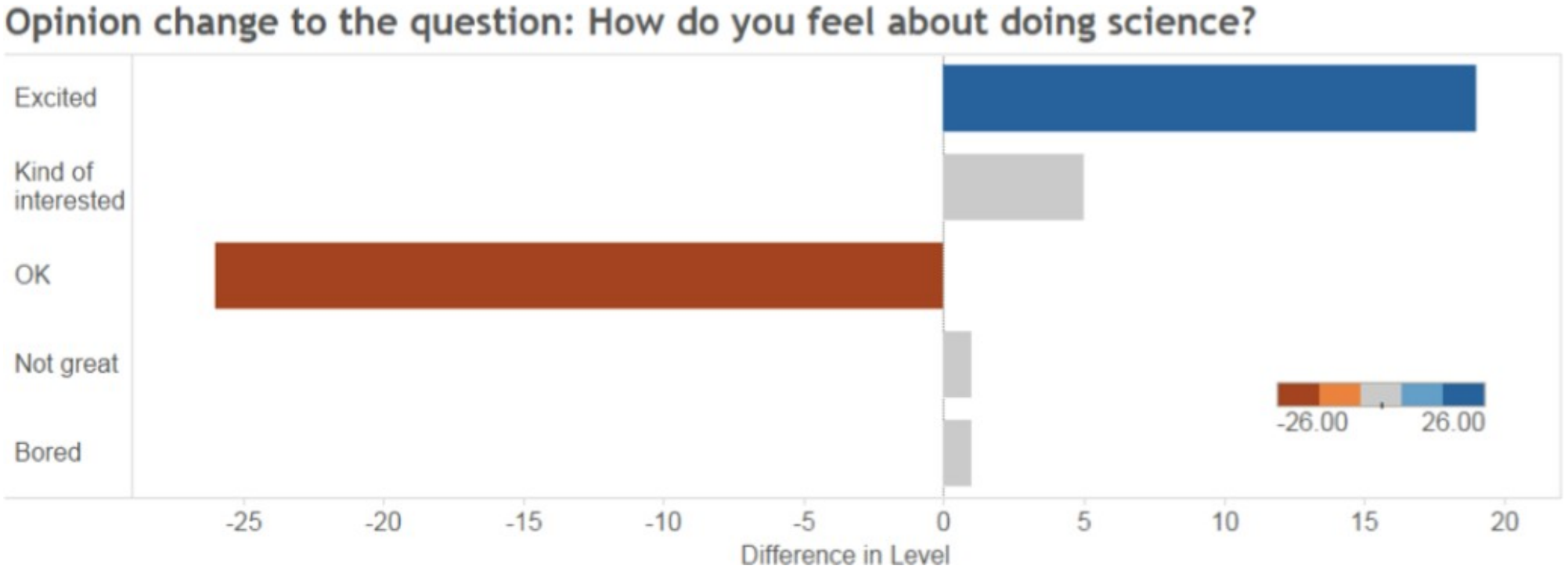
Sample visual 3

Clustered Column Chart



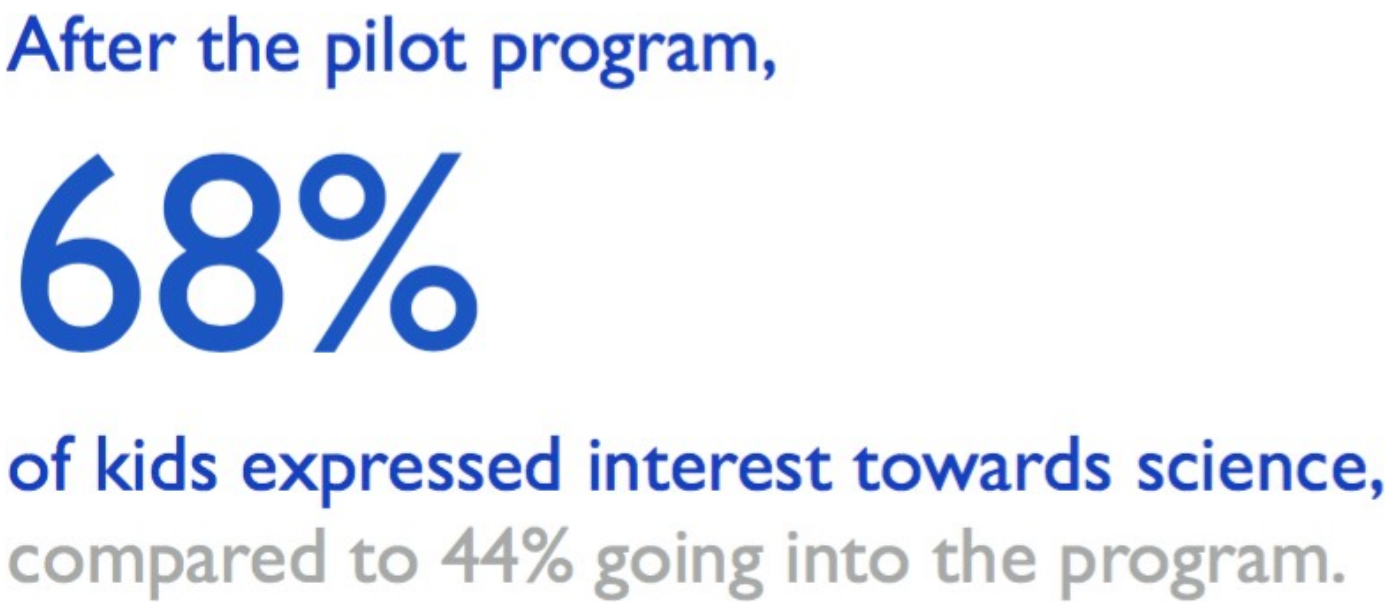
Source: Storytelling with Data – Cole Nussbaumer Knafflic

Design Activity
Sample visual 4: *Back-to-back bar chart*



Source: Storytelling with Data – Cole Nussbaumer Knafflic

Design Activity
Sample visual 5: *Impact Metric*



Source: Storytelling with Data – Cole Nussbaumer Knafflic

Know Your Purpose and Data

Know Your Purpose

Before you can begin to create stunning visualizations, you will need to make sense of your **data** by **finding the story** that speaks to your audience. Use the data to **illuminate that story** and the message you are trying to share, and you'll make it unforgettable.

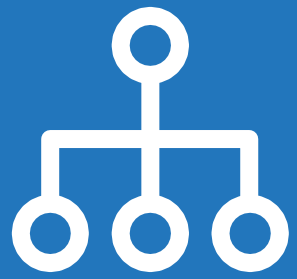
Know Your Data

It is important to model your data appropriately, before you explore it, in order to be able to answer your business questions correctly. **Data types** can be used **to model certain characteristics** of your data.

Numerical Data

1 2

AND



Dimensions

(Categorical)

Measures

Measures constitute **numerical data** that are **calculated** or **aggregated** – like the sum of Revenue, average Cost, or Profit-per-capita or non-numeric data that are counted.

Measures are objects that **represent calculations** and **aggregate functions** that are usually applied to **numeric data**.

Aggregating the object must make sense for the column to be a measure.

Sales Revenue is a measure but summing up product list prices isn't. That's a dimension. You can create measures from categories by counting their elements, for example, Number of Countries visited by our Customers.

What do measures represent?

Measures can represent **observations** in your data or **calculated values**.

How are they formatted?

Measures have an aggregation type associated with them. By default, BI Tools sets this type to **sum**. For example, if the chart includes Revenue by Country, and sum is associated with Revenue, the tool allows you to **customize the prefix or suffix** to indicate data such as, units of measures, like CAD, EUR, and USD.

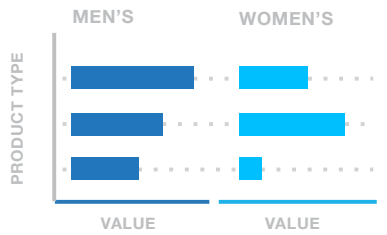
Dimensions

Dimensions constitute categorical data such as **year, product, country and salary range.**

What do they represent?

Categorical

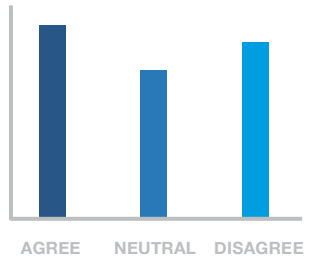
(Also called "nominal") for discrete values.



The dimension Product Type may include the values Men's Clothing and Women's Clothing.

Ordinal

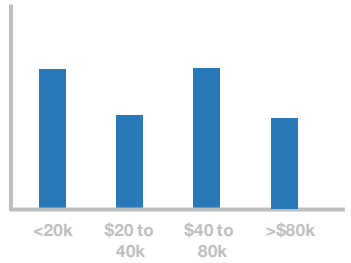
The dimension members have a set **default order**.



A dimension reflecting the outcome of a survey result may include the values Agree, Neutral, Disagree that have an implicit order.

Interval

Each value in the dimension represents **a range of values.**



The dimension Salary can be categorized into the following salary ranges: <$20k, $20 - 40k, $40 - 80k, >$80k

Craft Your Message

By **exploring** your **data** you now have a **better sense of what story** to tell your audience.

It is time to craft that message and discover which viz best articulates your information.

Keep these questions in mind:

1.

What is your overall goal of your data analysis?

2.

Who is this message intended for? What do you know about your audience?

Then ask yourself:

1. What **questions** do you want to **answer** with your data?
2. What kind of **relationships exist in your data**?
   What are the best **techniques** for **displaying** these?
   Do you need a chart (overview), a table (details), or maybe both, to convey your message?
3. Can you **highlight specific data points** to better get your message across?
4. How can you **incorporate a summary** of your message in your chart titles to emphasize on your overall message?

Know Your Audience

Get to know your audience then use **precognitive attributes** to create great data visualizations that resonate with them. Precognitive attributes mean the **image is being processed without any conscious effort**. Communicating in this way means there is no need for explanation on top of the visualizations. It is also important to note that just because you have good visualizations that doesn't necessarily mean you have a good visual story. Reward your audience with the experience and knowledge that led them to your story in the first place.

Every piece needs to be pulled together to create a cohesive story with a beginning, middle, and end. Entertain them.

Storytelling Assets

Selecting the Right Visualizations

Change Over Time

Shows how a measure changes over time, and allows the user to highlight temporal trends

- Line Chart: Highlights potential trends in data
- Bar Chart: Highlights comparison between individual values

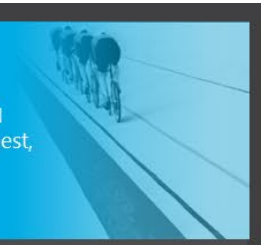
Comparison

Shows the comparison of categorical values, where the data does not have any intrinsic order, for example, a list of products

- Bar Chart: used for comparing categorical values
- Trellis: uses multiple views to show different partitions of a dataset

Ranking

Shows the top or bottom N values to emphasize the largest, or smallest values



- Bar Chart: shows categorical values in decreasing or increasing order

Part-To-Whole

Shows how the categories contribute to the whole value

- Bar Chart: set to % scale
- Pie Chart: Compares percentage values
- Stacked Bar Chart: shows overall measure total

Distribution

Shows how a measure is spread across its domain

- Histogram: Column Chart showing the count of binned measure values
- Box Plot: shows distributions for different categorical values
- Heat and Tree Map: shows the distribution of measure values

Correlation

Shows, whether there is a potential correlation between two measures

- Scatter Plot: highlights potential correlation of two measures
- Trellis: uses multiple views to show different partitions of a dataset

Overview

Shows the exact values in table format

- Table: highlights exact values

Geographical Information and Maps

Shows the geographical distribution of measure values

- Choropleth Chart: highlights geographical data by colouring geographical areas according to their measure values
- Geo Bubble Chart: highlights geographical data by showing them as bubbles on a map

Change Over Time

Shows how a measure changes over time and allows the user to highlight temporal trends.

Focus Areas

Line and Area Chart

Bar Chart

Line and Area Chart

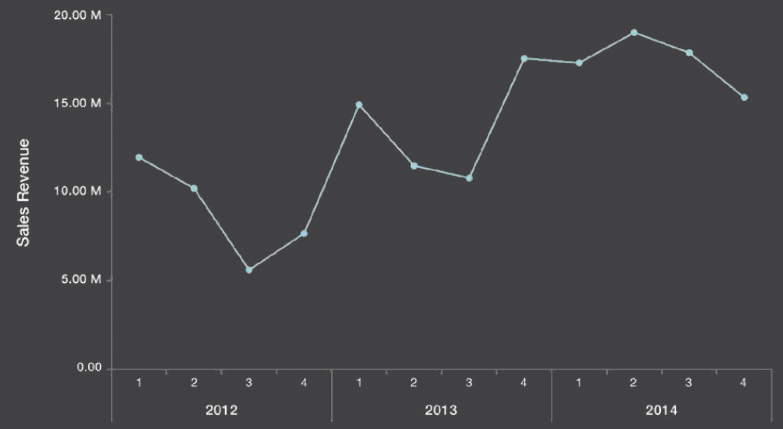
The Line Chart **displays measures over a time period.**

Line Charts are used frequently **to show trends** and **relationships** between them. The Y-Axis always shows a measure value, and the X-Axis denotes a time dimension such as Month, Quarter, or Year.

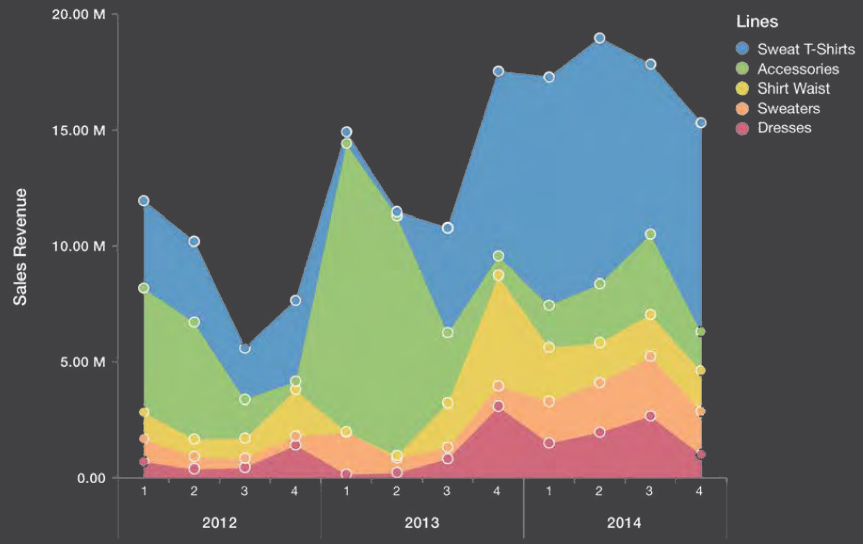
Used for

- Trends
- Data over time
- Temporal patterns and correlation
- Period-over-period

Sales revenue in 2012 - 2014



The impact of different product lines by sales revenue in 2012 - 2014



Suggestions

1.
Create a time hierarchy to allow drilling up or down to Days, Months, and Years

2.
Add a moving average line to smooth the trend over time

3.
Add a forecast or linear regression to emphasize current or future trends

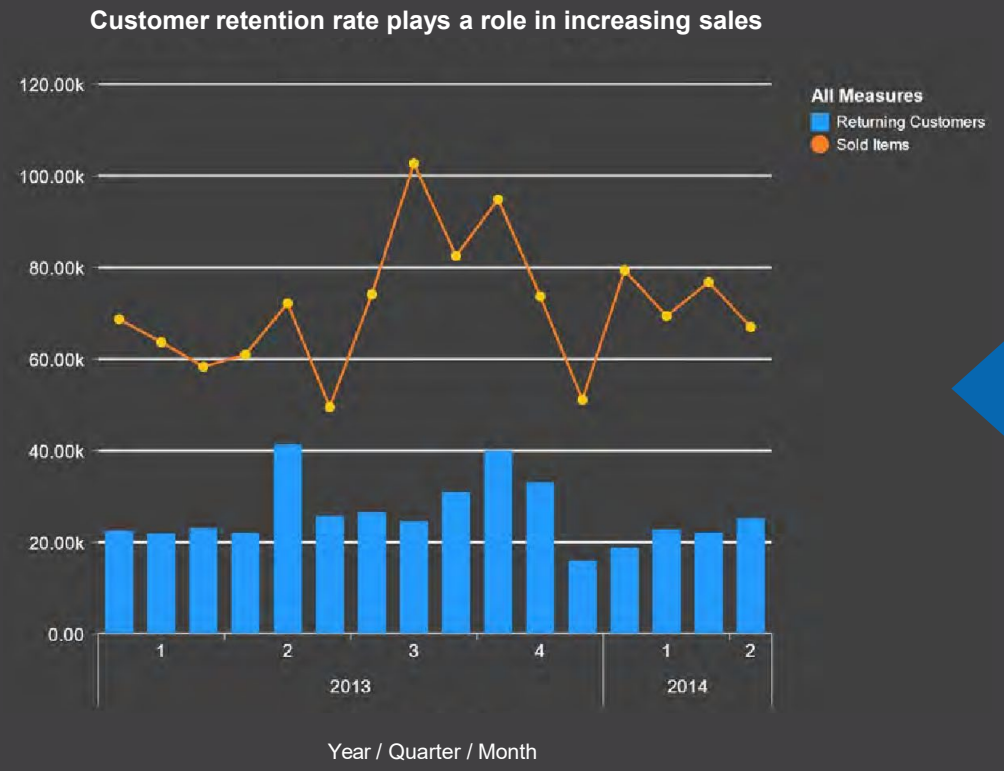
4.
Consider an Area Chart for showing cumulative totals

Column Line Chart

The Column Line Chart is a **combination of a Line Chart and Column Chart.** This chart type displays **one measure as a column** and **a secondary measure as a line**. The two measures are displayed over a **Time Dimension** which may include Years, Quarters, or Months. This chart is great for showing the relationship between two measures over a period of time such as Gross Margin and Sales Revenue, or Net Income after Tax and Tax Rates.

Used for

- Trends
- Data over time
- Temporal patterns and correlation



Suggestions

1.
Use this chart type to show two trends of different types (for example, Returning Customers and Sold items) over time

2.
Other options for showing change over time include Bar Charts or Tables.

Comparison

Shows the **comparison of categorical values**, where the data does not have any intrinsic order, for example, a list of products.

Focus Areas

Bar Chart

Trellis

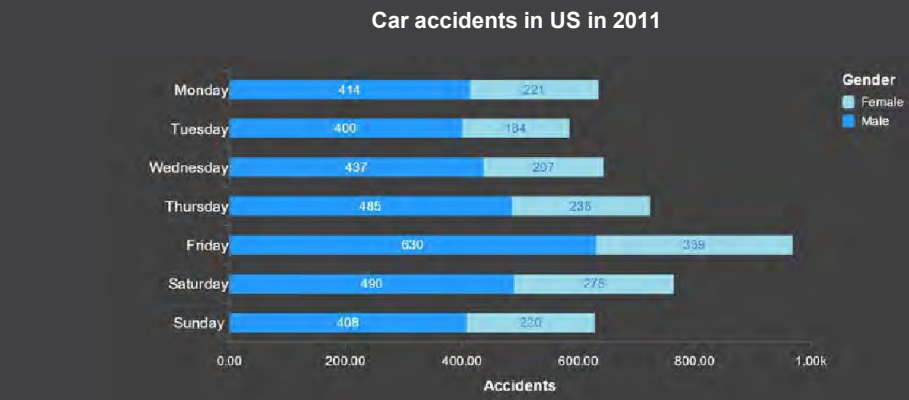
Bar Chart & Stacked Bar Chart

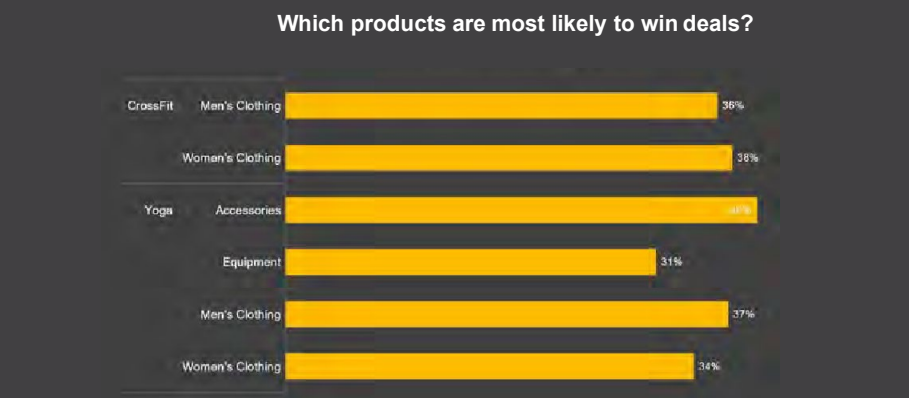
Bar Charts are probably the most frequently used chart type. Focus the attention of your audience to important details by:

- **Ranking data** from **largest to smallest** or **vice versa**
- Filtering out data that isn't important for your message
- Grouping data by combining values in a chart – if there are too many categories, you can group less relevant categorical values together into an Other group (for example, "Other Clothing")

Used for

- Comparing different categorical values





Suggestions

1.
Use data labels, such as Figure 5, to improve the readability of data values

2.
Customize hierarchies to allow drilling from a high-level overview to more specific details; users easily drill up and down

3.
Use Color to clearly differentiate separate categorical values in your dimension

Ranking

Shows the **top or bottom N values** to **emphasize the largest, or smallest values.**

Focus Area

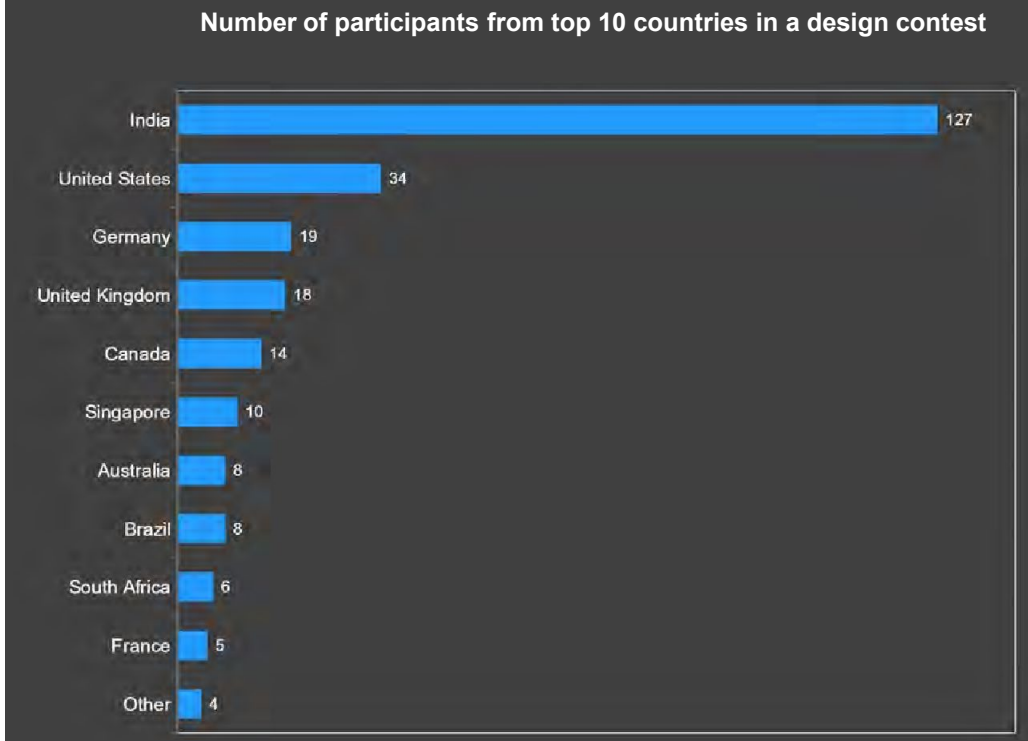
Bar Chart

Ranking

The Ranking Chart allows the user **to sort and filter data based on their importance**. For example, we may want to sort Countries based on their Number of Participants.

Used for

- Emphasizing on top or bottom values in a chart



Suggestions

1.

Often categorical values (in this case Countries) that contribute less to the overall measure value might be filtered out or grouped together in another category.

Part-to-Whole

Shows how the **categories contribute to the whole value.**

Focus Areas

Bar Chart

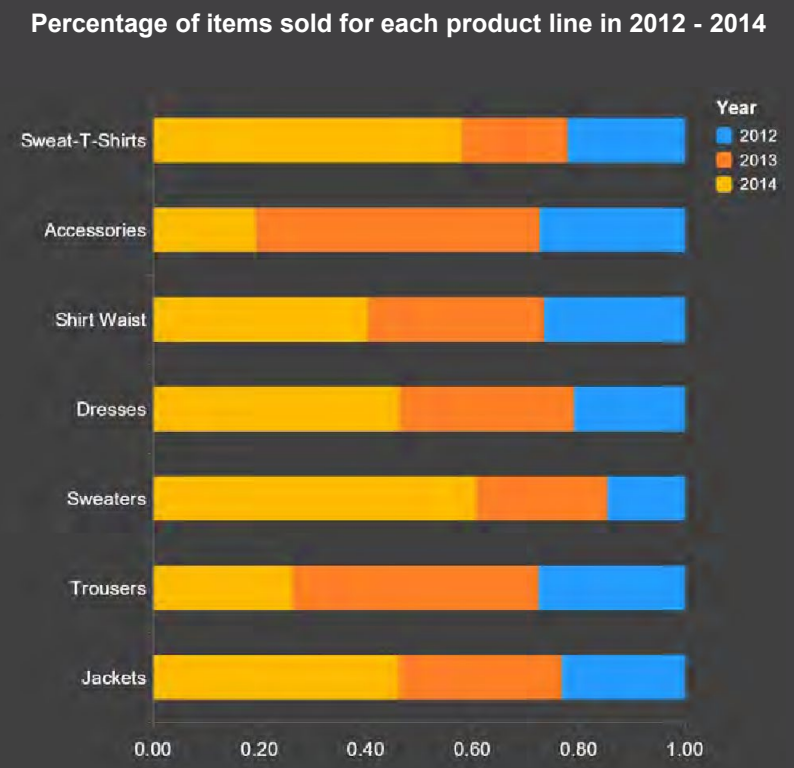
Stacked Bar Chart

Pie Chart

Part-to-Whole

Used for

A Part-to-Whole relationship shows **how measure values that make up the whole of something** (for example, Number of containers sold) **compare to one another** and how they each compare to the whole.



Suggestions

1.
You can use stacked or side-by-side bars to compare different hierarchy levels (Country Region) or classifications (Men's Clothing, Women's Clothing).

2.
You can use a 100% Stacked Bar Chart (or Marimekko Chart) to show the portion that each segment makes up in a category.

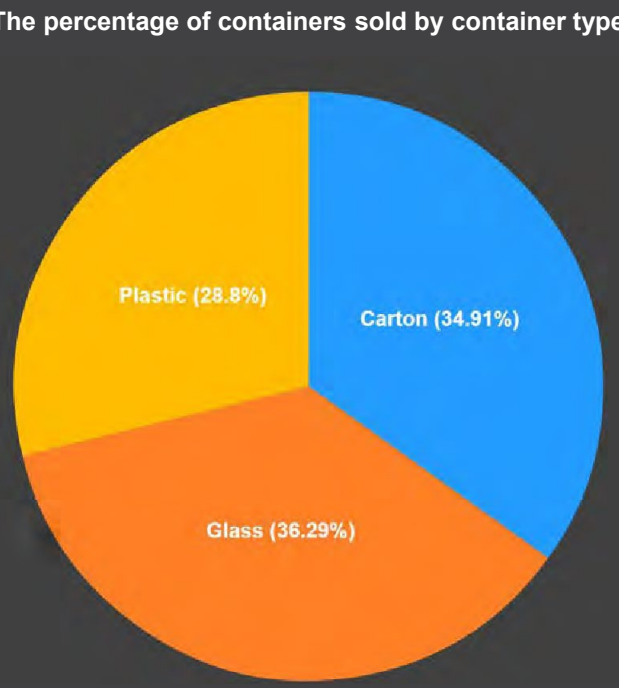
3.
In addition to Stacked Bar Charts and Marimekko Charts, other charts (such as pie, ring, and funnel charts) can be used to show Part-to-Whole relationships.

Pie, Ring, and Funnel Charts

Pie, Ring (Donut), and Funnel Charts are used to **discern part-to-whole comparisons to either highlight a portion of the data** or to **compare values for different categorical** values. These chart types are generally **not recommended if they include too many segments**, as the viewer will have a difficult time differentiating between too many different colors.

Used for

- **Comparing percentage** values in proportion to the whole



Suggestions

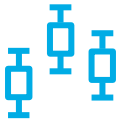
1. Limits use of Pie Charts to a small number of slices (no more than 7 slices)
2. Consider showing data labels for ease of reading
3. Highlight only the most important slice if possible
4. Compare with using a bar chart or ring (donut) chart – the viewer is more likely to perceive the length of a bar over the size of angular slices

Distribution

Shows how a measure is spread across its domain.

Focus Areas

Histogram



Box Plot

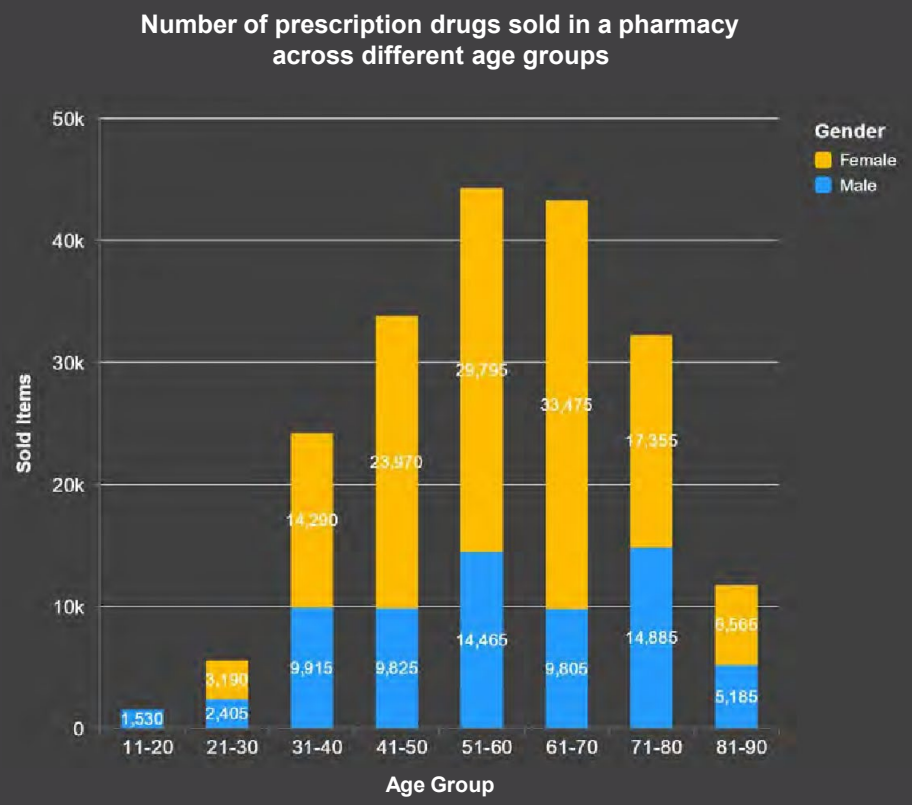
Heat and Tree Map

Histogram and Binning

A Histogram is a type of Column Chart that shows the **distribution of measure values,** for example, Number of Items Sold. Instead of showing each measure value directly, in a histogram, values are binned first. For example, in Figure 9, instead of creating one column per Age, we binned the values first into the age ranges [11-20], [21-30].....[80-90]. This allowed us to show the distribution of Number of Prescription Drugs Sold in an audience friendly way.

Used for

- Distribution of measure values, identifying data issues including outliers



Suggestions

1.

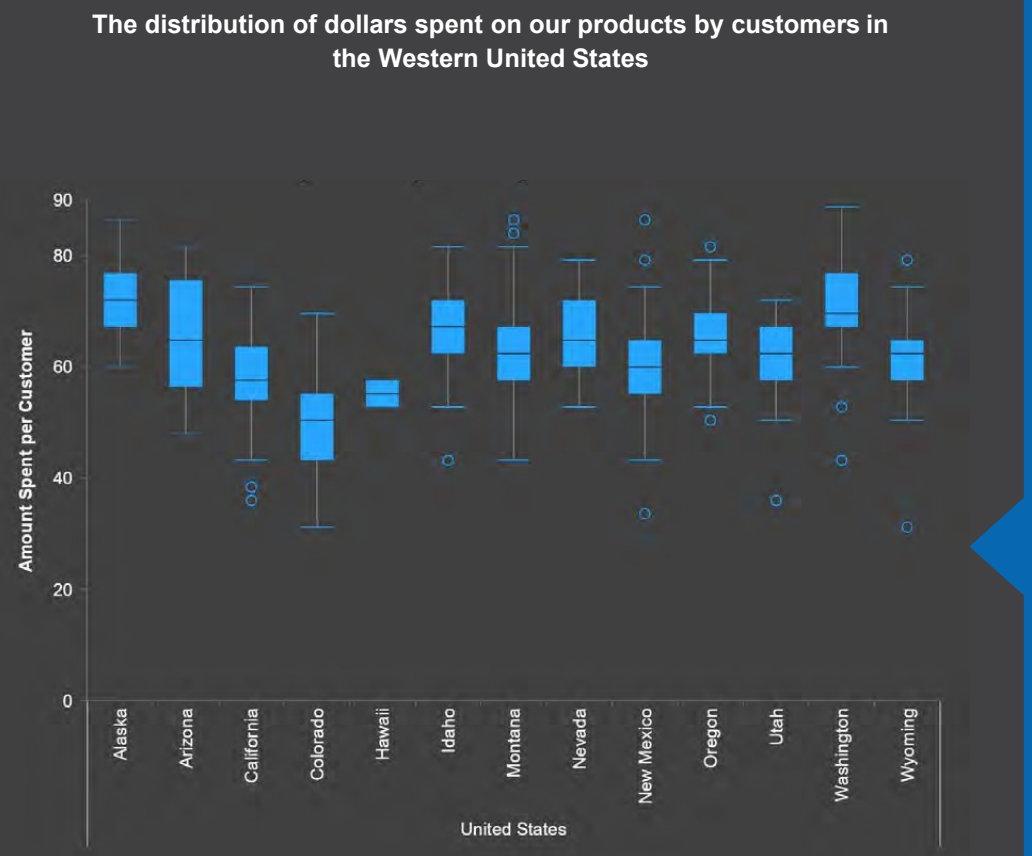
Create bins or ranges of numbers to count the number of occurrences within your data. In SAP Lumira, this can be done either in the Prepare Room using the Group by Range functionality or in the Visualize Room using a Calculated Dimension.

Box Plot

A Box Plot visually **displays statistical distribution of a measure within a dataset.** It is often used to also show the range in values for each categorical value. Boxplots show the **minimum** and **maximum** values, as well as the **median** and the 25th and 75th quartile. Outliers are visually represented by dots.

Used for

- Comparison, distribution of values, identifying outliers



Suggestions

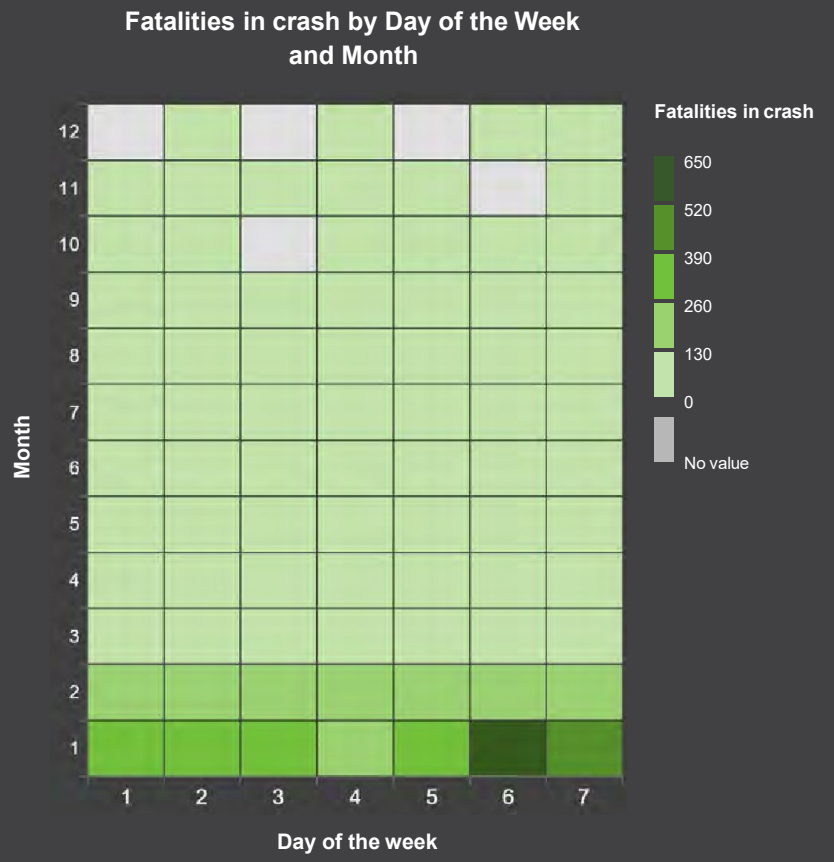
1. Compare data distribution for several categorical values
2. Show distribution of medians in data
3. Include a reference line for the overall median in your data

Heat Map and Tree Map

In a Heat Map the **categorical values** are contained in a matrix of tiles; based on a single measure, these tiles have different shades. In contrast, in a Tree Map, two measures are considered. Larger values are represented by larger tiles and darker shades.

Used for

- Showing the distribution of measure values



Suggestions

1.

Only use this if the resulting Heat Map shows visibly different color intensities (it will confuse the viewer if the heat map segments are of similar color intensities).

Correlation

Shows whether there is a **potential correlation between two measures.**

Focus Areas

Scatter Plot

Trellis

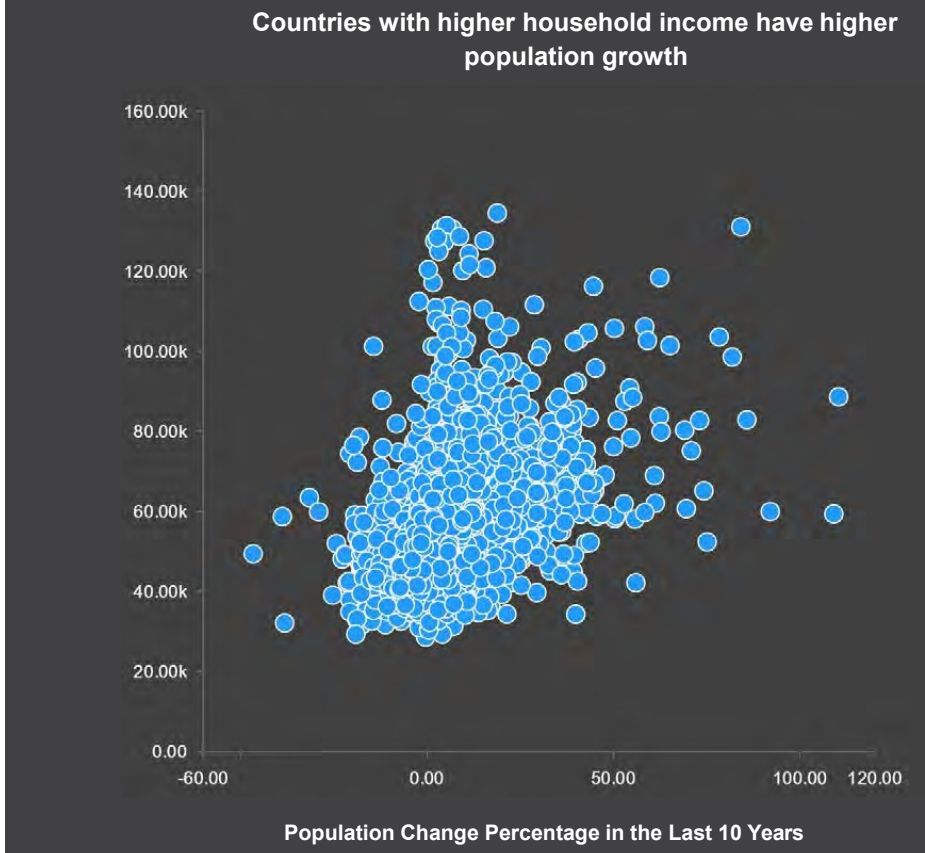
Scatter Plot

Additional Chart Types Used For Showing Correlation:

- The Scatter Matrix shows several Scatter Plots in a Trellis layout in order to compare several Scatter Plots in one chart.
- A Bubble Chart is **similar to a Scatter Plot** but allows visualization of a third measure. The **size** of the bubbles indicates this third measure. **The larger the measure is, the larger the bubble**.

Used for

- Showing the correlation of two measures



Suggestions

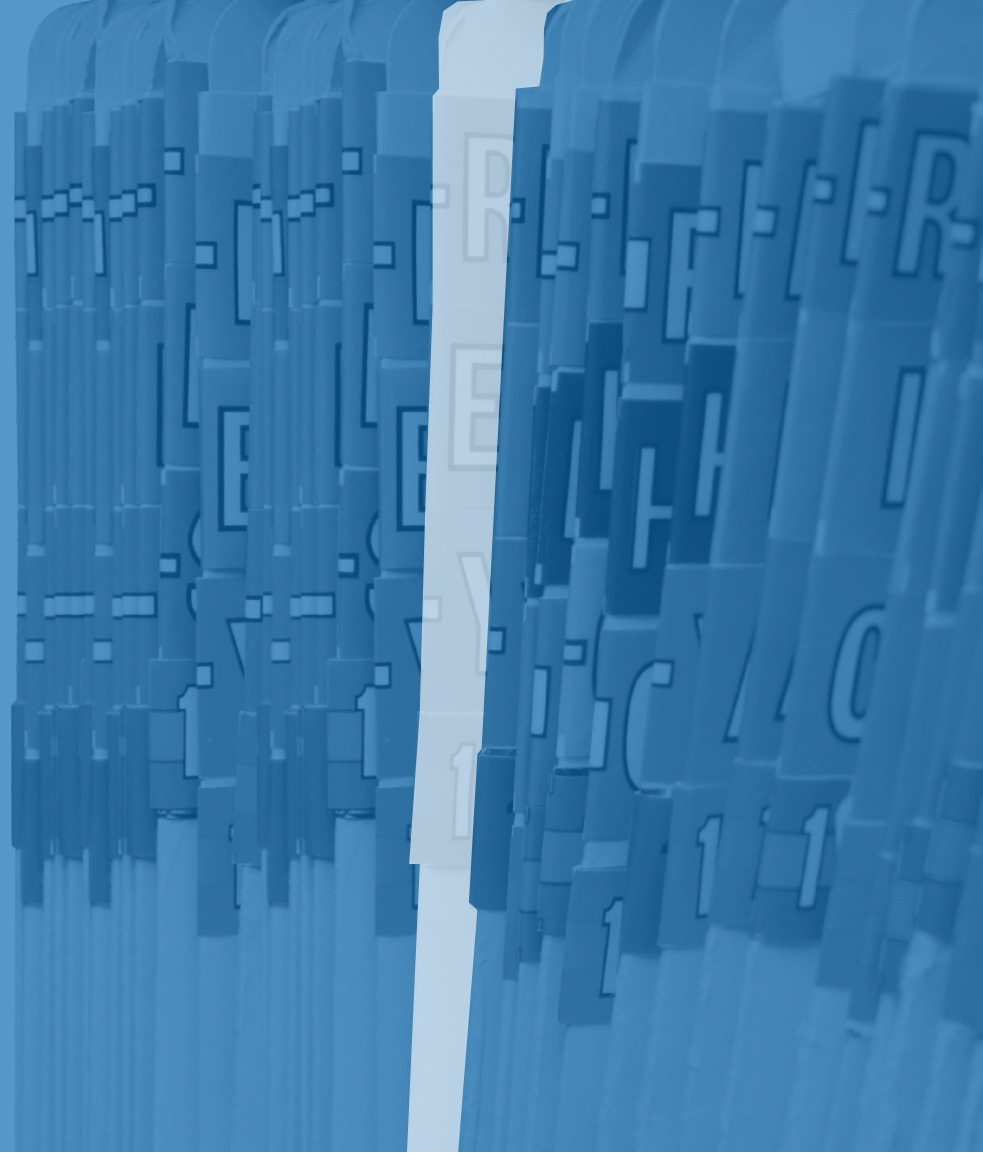
1.
Use the color to show groups of points, but limit the number of colors used; too many colors or shapes will impact the readability of a chart

2.
Keep the aspect ratio square

3.
Create a Geo Hierarchy on top of location data (for example, States, Cities) to enable drilling up to higher levels of geographical detail

Overview

Shows the exact **values in table format**.



Focus Area

Table

Table

A table is an arrangement of data in rows and columns, or possibly in a more complex structure.

The elements of a table may be grouped, segmented, or arranged in many different ways, and even nested recursively.

Used for

- Show multiple measures in one or two categories or hierarchies

Drug and alcohol usage had a significant impact on fatal injuries in the United States in 2011

Injury Severity	Gender	Measures: Accidents	Accidents Involving Drugs	Average Blood Alcohol Content
Fatal Injury	Female	716		0.05
	Male	1,603	108	0.08
Incapacitating Injury	Female	227	8	0.02
	Male	356	10	0.05
Non-incapacitating Evident In	Female	281	7	0.02
	Male	403	13	0.03

Suggestions

1.
Best for showing exact values.

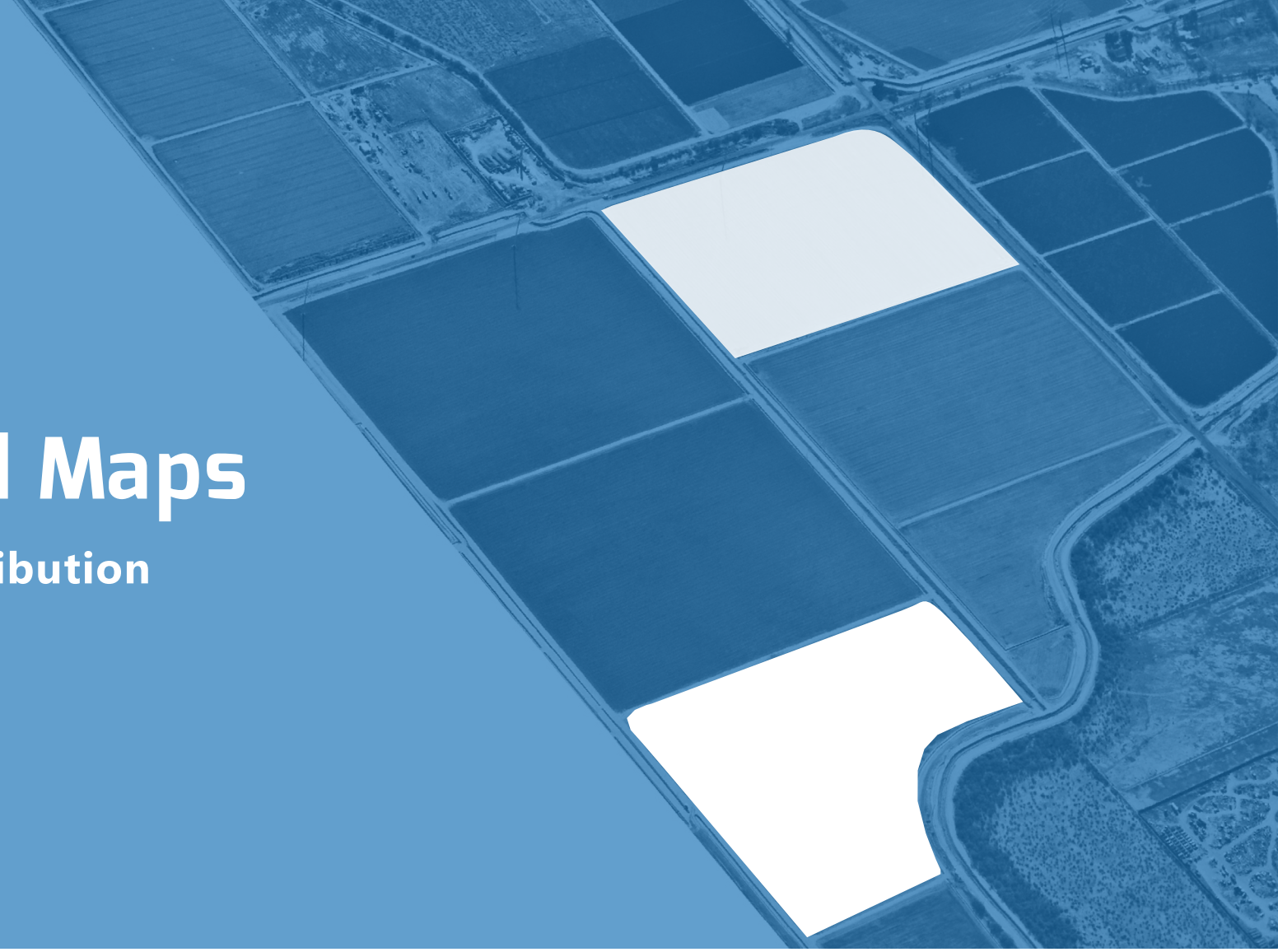
2
Often charts and Tables are shown on the same page, as they emphasize different aspects.

3.
Highlight key information with the Conditional Formatting feature.

4.
Setting the correct precision (number formatting) for measures included in a Table is paramount in order to not overload the user.

Geographical Information and Maps

Shows the **geographical distribution of measure values.**



Focus Areas

Choropleth Chart

Geo Bubble Chart

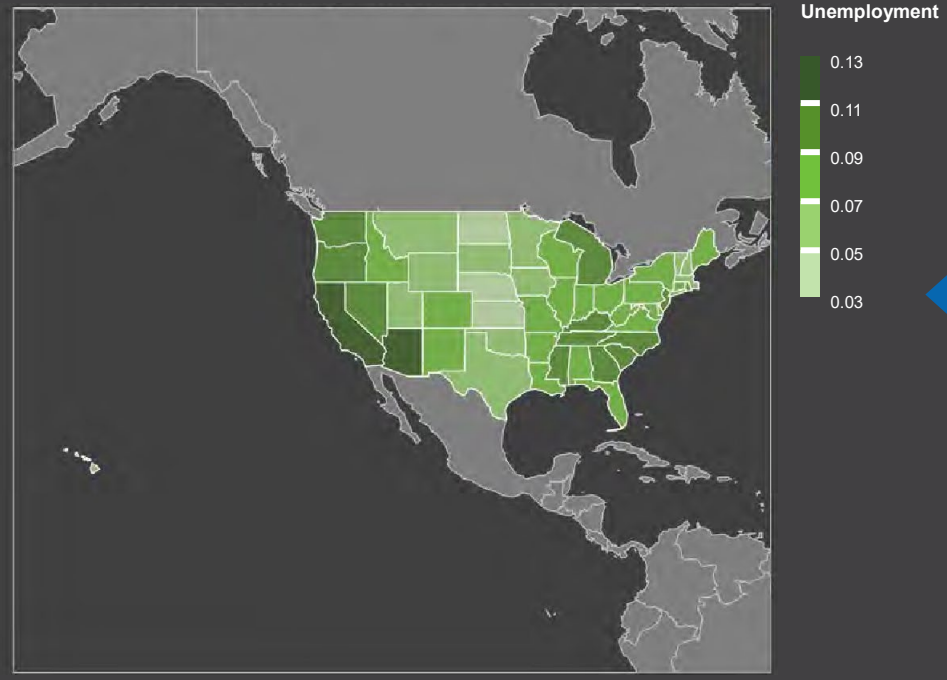
Choropleth Map

A Choropleth Map uses differences in shading, coloring, or the placing of symbols within predefined regions to indicate measure values in those areas.

Used for

- Supports location-based comparisons of standardized data such as Rates, Densities, or Percentages

Unemployment percentage across different states in the United States in 2014



Suggestions

1.
Use the Choropleth Map for locations of similar size, as the size of the area coloured may overemphasize larger areas (for example, Canada covers a much larger area than Japan despite being much smaller in terms of population)

2.
Make sure your measure values are normalized by the geographic properties, for example, by the population of a geographic area

3.
Remember that the granularity of your regions (counties, for example) will impact the signal (aggregated measure values) from your data.

Geo Bubble Chart

The Geo Bubble Chart shows **measure values in the form of bubbles on a map**. The larger the measure, the larger will be the bubble on the map.

Used for

- Viewing a measure by Country, Region, or City; comparing
- measures across different geographic areas

Voyages ending in Americas experience lower delays

Suggestions

1. Use to show values on a map and to create an animation over time

2. Use Geo Bubble or Pie Charts on maps to show measure values if the relative size of the underlying regions cannot be compared

Protips

A great visual design standard will **accelerate understanding** and **consumptions of your data**. It's that simple. For your business to reap the benefits of data visualizations, organizations need to create a **visual design standard** that **incorporates best practices.**

The International Business Communications Standards (IBCS), is a non-profit organization that has established a rather comprehensive and detailed visual standard for designing both reports and presentations. It is highly recommended that anyone who develops reports, either as a data professional or business analyst, should peruse both the IBCS Web site (www.ibcs-a.org) and Rolf Hichert's consulting Web site (www.hichert.com).

1.

Less is more. Make every pixel and word count.

2.

Avoid decorative use of graphics.

3.

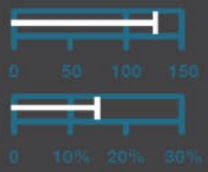
Avoid three-dimensional chart types.

4.

Avoid pie charts.

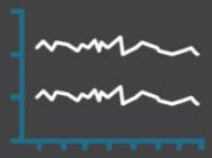
5.

Start bar charts at zero.



6.

Use bullet graphs instead of gauges to save space.



7.

Use sparklines to show trends on the X-axis.

8.

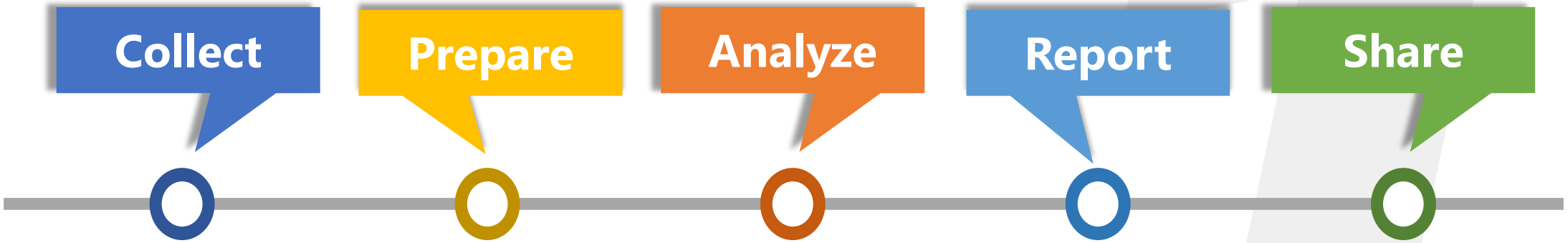
Show time going from left to right on the X-axis.

9.

Use color only to highlight or accentuate meaning.

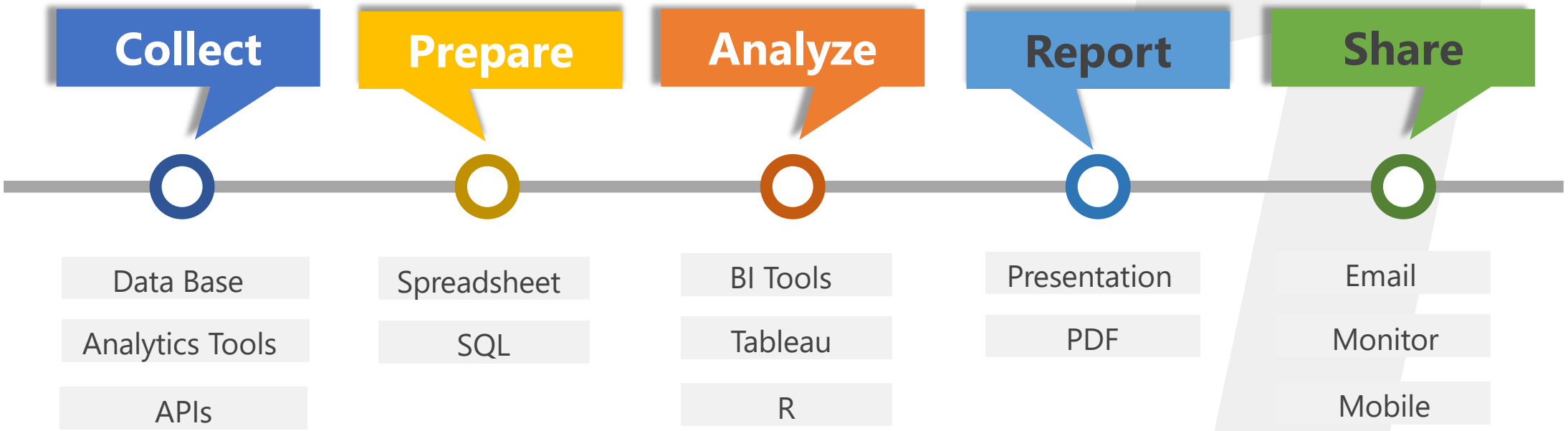
Google Data Studio + Demo

Common Data Workflow: Challenges



Limitations in any stage cause users to exit and use different tools.

Common Data Workflow: Challenges



Customers use many tools. It's complex, expensive, and difficult to manage

Common Data Workflow: Challenges

Data Studio Simplifies the Analysis Workflow

DATA SOURCES

Access Data

- Connect & Import Data
- Semantic Classification
- Define Transforms

ANALYSIS

Explore Data

- Visual Query
- Drag/Drop Data
- Dynamic Filters

PRESENT & SHARE

Communicate Data

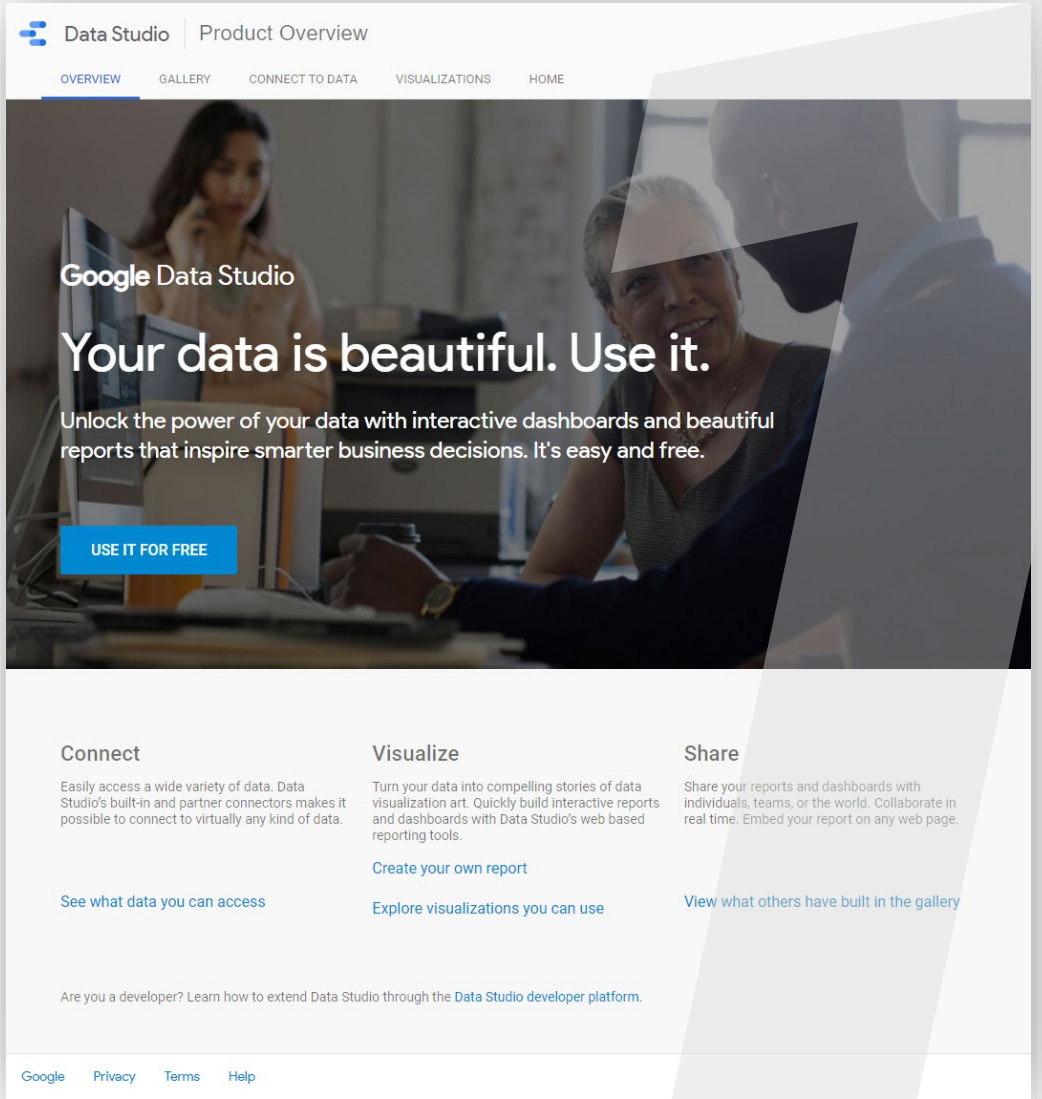
- Modern WYSIWYG Editor
- Charts/Shapes/Styles
- Interactive Controls
- Drive Sharing & Collaboration

What is Google Data Studio?

- Initially Google data studio was first launched as part of **Google Analytics Suite 360 in June 2, 2016**
- It was **premium version** and was only available within the US.
- Later, on **May 24,2016**, Google announced a **free version of data studio** for individuals and small teams.

Google Data Studio

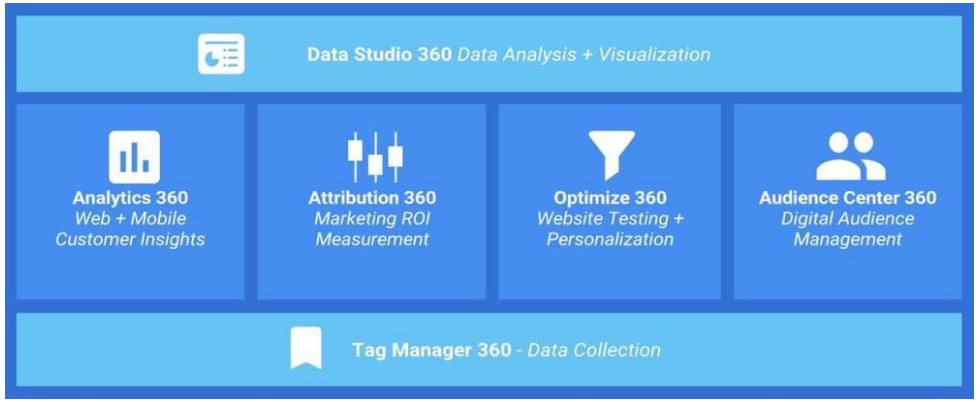
Google Data Studio



Piece of puzzle

Under the hood of **Google Analytics suite 360**

- Even though its capable of doing a lots of things, Google Data Studio is just single puzzle part of the whole big picture.
- Data Studio part of Google Analytics 360 Suite, a platform that will help you evaluate the full customer journey and drive results. The Suite is comprised of 6 products, as schematized below:



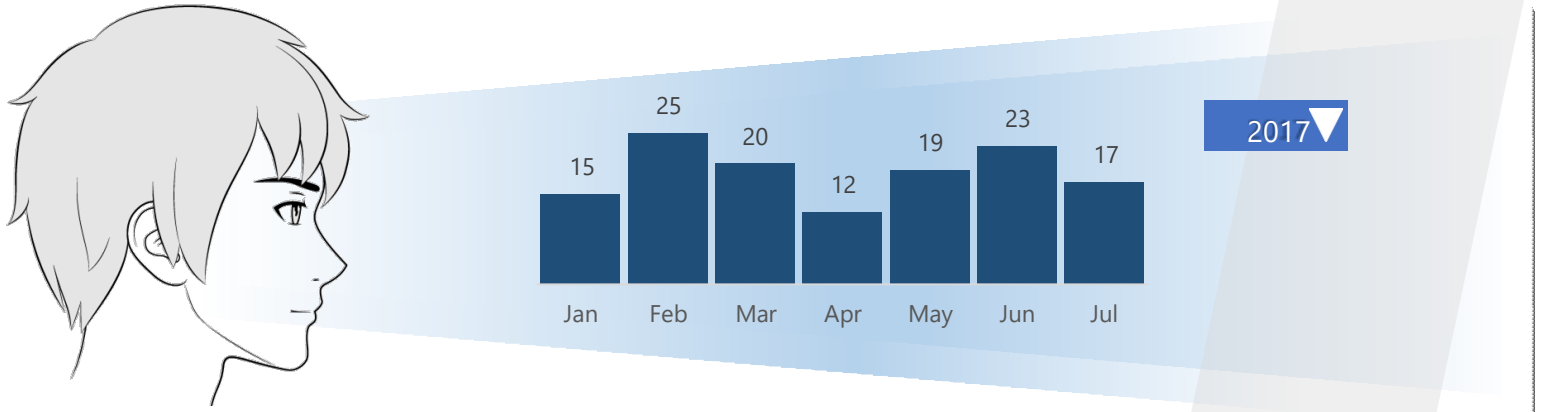
What is Google Data Studio?

- is a **reporting tool** which turn data into **informative dashboards and reports** that are easy to read, easy to share, and fully customizable.
- able to **automatically update data in real time**.
- gives freedom to multiple people to **work on same report in real time.**
- Helps business be more data-driven by making access to data and analysis easier
    - Available for Totally Free
    - Unlimited Reports
    - Free 2 GB Data Import

Why use Google Studio ?

- **90%** of the information a person **perceive is through sight**
- **70%** of the **sensory receptors** are located in the **eyes**
- About **50%** of the human brain neurons are involved in the **processing of visual information**
- Human **productivity** is **17%** higher when working with **visual information.**
- A person **understand visual information 60,000 times faster** comparison to the same **information as text**

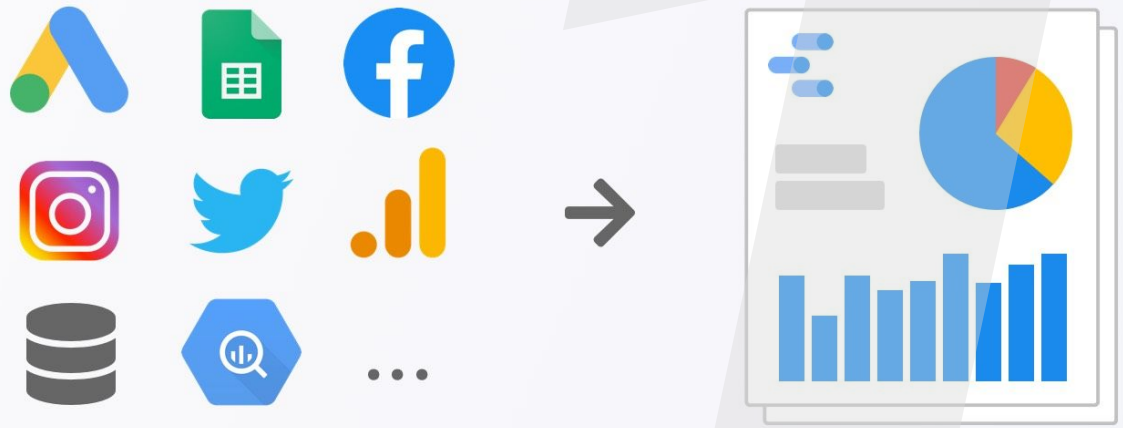
Jan	15
Feb	25
Mar	20
Apr	12
May	19
Jun	23
Jul	17



What you can do with Data Studio.

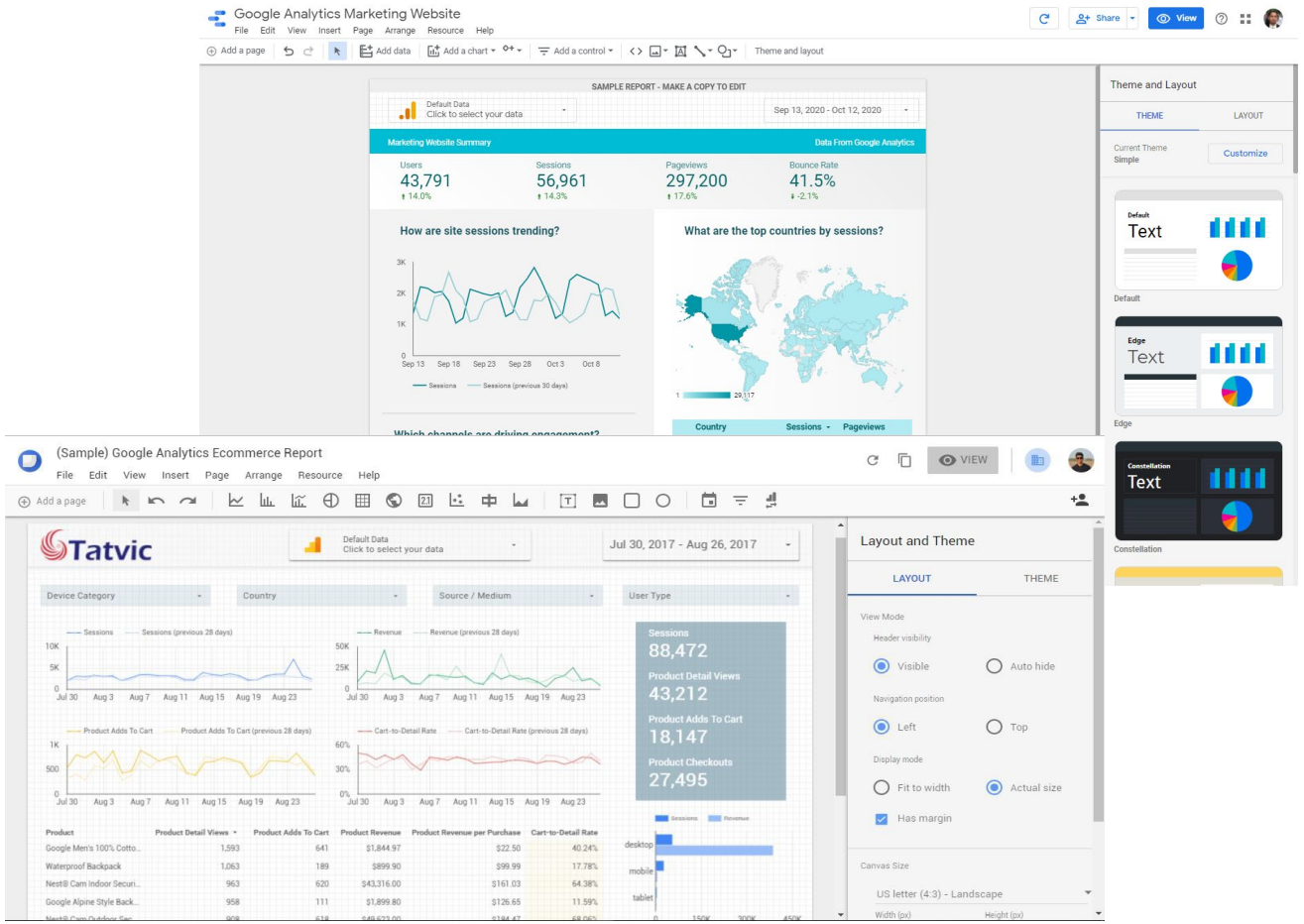
Tell your story in data:

- Visualize your data through **highly configurable charts** and **tables.**
- Easily **connect** to a **variety of data sources**.
- **Share** your insights with your team or with the world.
- **Collaborate** on reports with your team.
- Speed up your report creation process with **built-in sample reports**.



Google Data Studio Features

- **Connecting Data** from different sources
- Creating own **Dimension**, **Calculation of Metrix**
- Using **Filter** controls and date ranges
- How to use **score cards**, **Maps** and **Tables**
- How to use **Charts, Bullet Chart** and **Combo Chart**
- Setting the **layout**, **Pages** and **style** of the report
- **Publishing** and Embeding the report
- Creating Segments and Filters
- Grouping selected data and date ranges



What you need to use Data Studio?

Prerequisites to **view**, **create**, and **share** Data Studio files.

What you need to view reports

- To view a Data Studio report, all you need is a **browser** and an **internet connection**.
  Note: A Google account is not required to view reports.

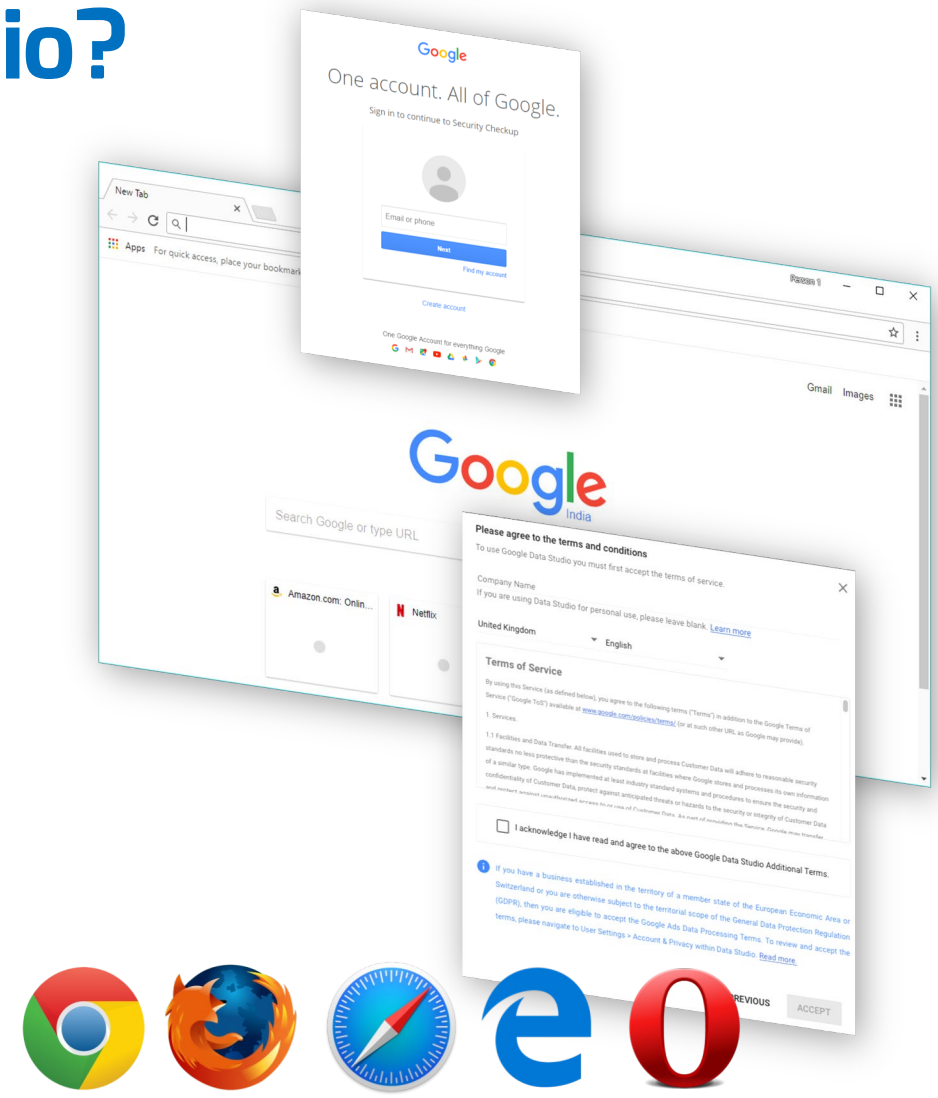
What you need to create and edit reports and data sources

To create or edit a file in Data Studio, you must:

- be signed into a **Google account**
- Learn how to map a non-Gmail address to a Google account.
- be in one of the **supported countries** (see below)
- **accept** the Data Studio **Terms of Service and Policies**
- If you are a G Suite user, your administrator may need to enable Data Studio.

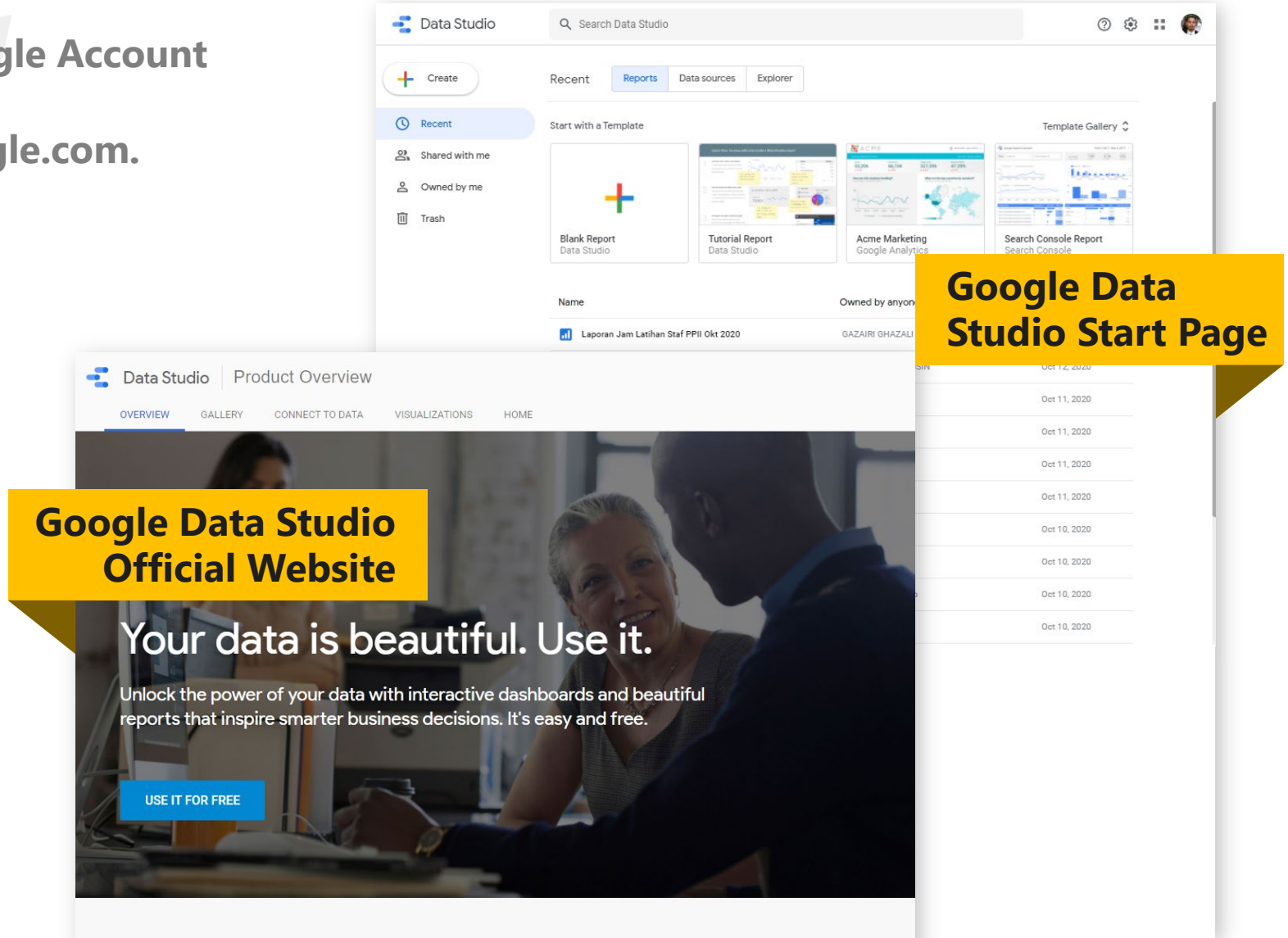
Supported browsers

- Tested on the Chrome, Firefox, and Safari browsers.
  Note: Editing reports and data sources on other browsers is not recommended at this time.



Starting with Google Studio

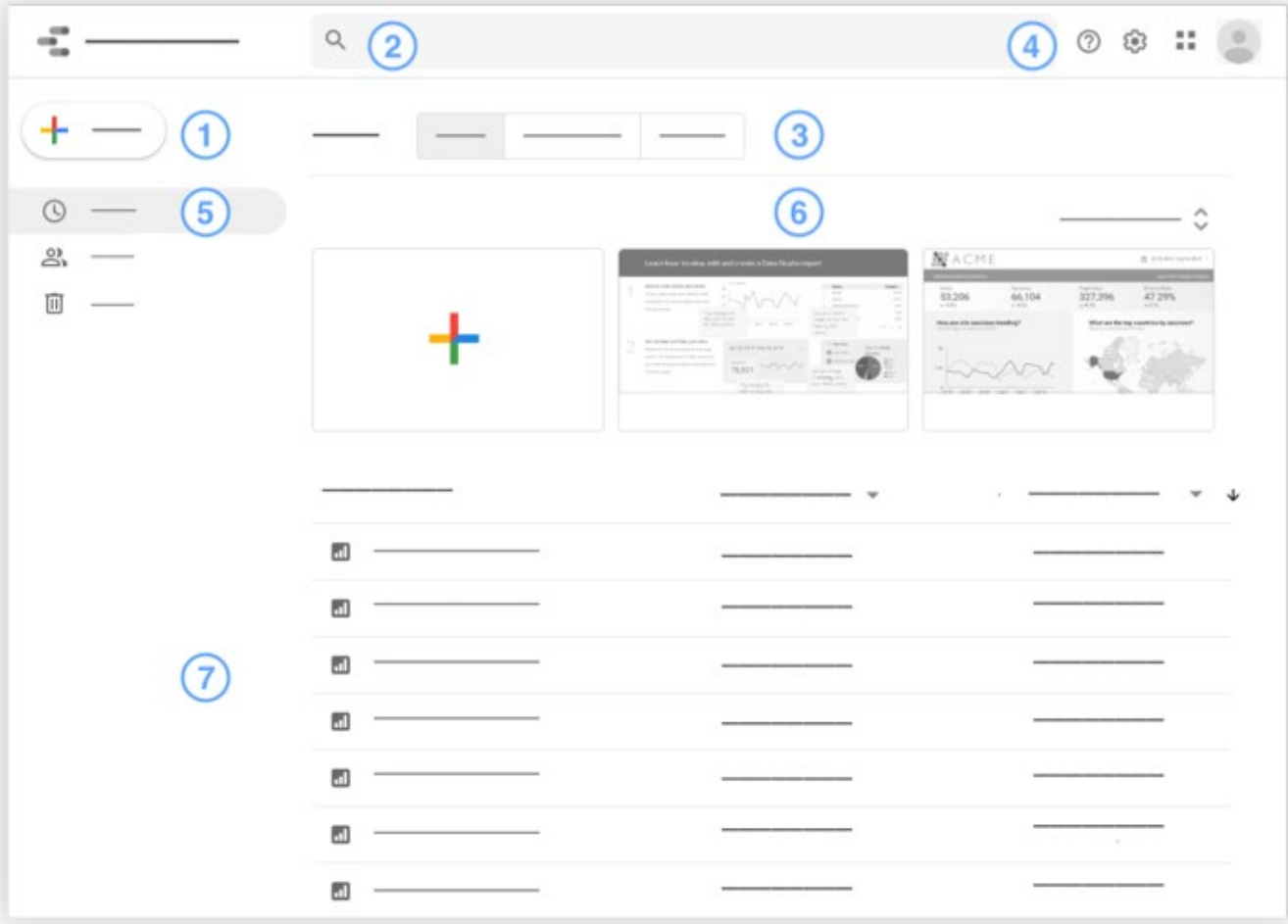
- First, log in with the **Google Account**
- Login at **https://datastudio.google.com.**



Find your way around Data Studio

The home page is where you create and access all your Data Studio files.

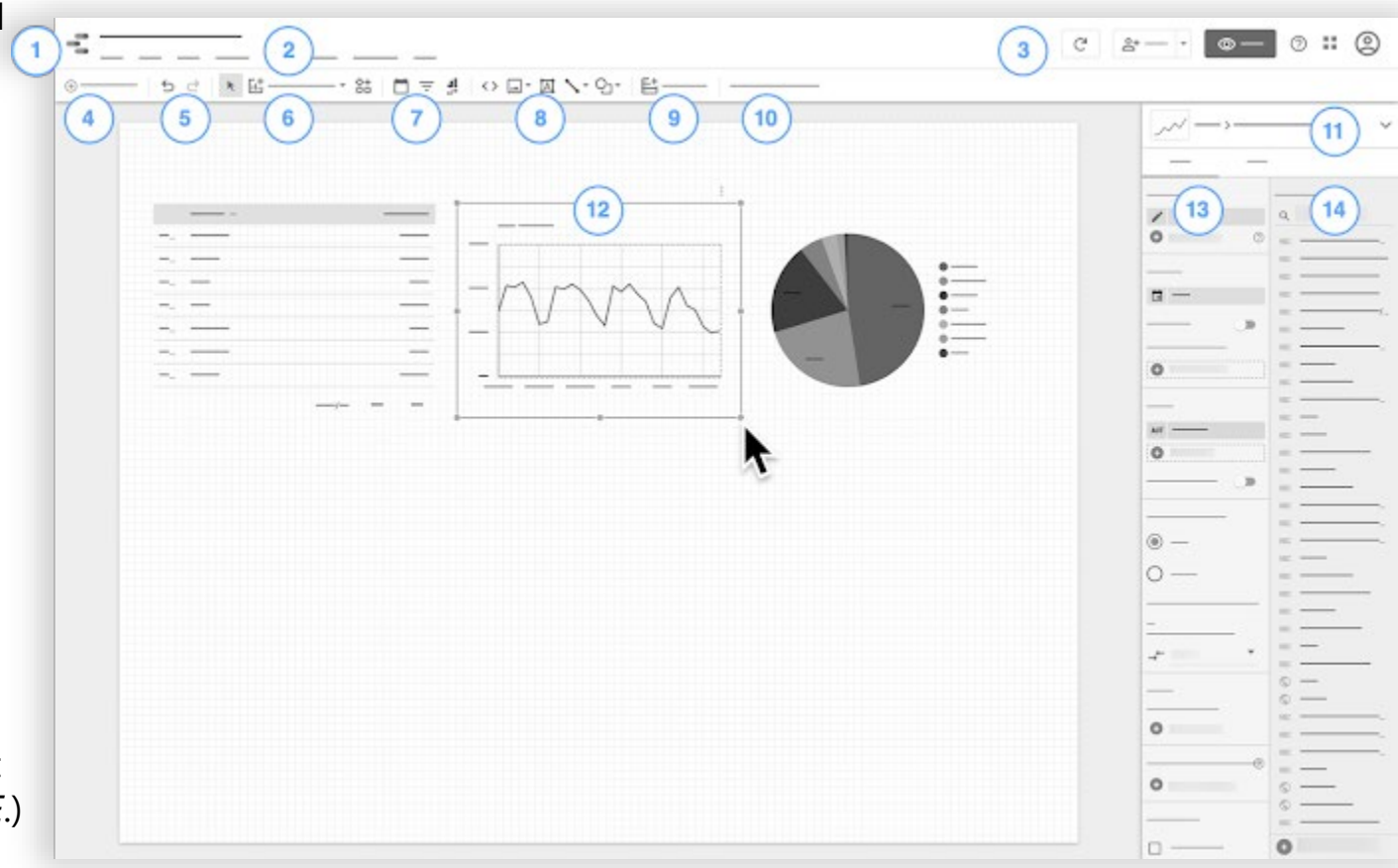
1. **Create** a new file.
2. **Search**. Find files quickly .
3. **File type tabs**. Switch between Reports, Data Sources, and the Explorer.
4. **Options**:
   - ***Help*** *and feedback.*
   - ***User settings***.
   - ***Google Marketing Platform*** *product selector (GMPusers only)* .
   - ***Manage*** *your* ***Google account*** .
5. **Filter the file list to show recent files**, those owned by you, those shared with you, and files in the trash.
6. **Report samples and templates**. Start with a blank report, or customize a fully functional template.
7. **File list.** Click a file to view it.
   - On the right, use a file's overflow menu to share, rename, or remove it .
   - Sort files by name, owner, or date.



Find your way around The Report Editor

To edit a report, locate it in the file list, view the file, then in the upper right, click EDIT.

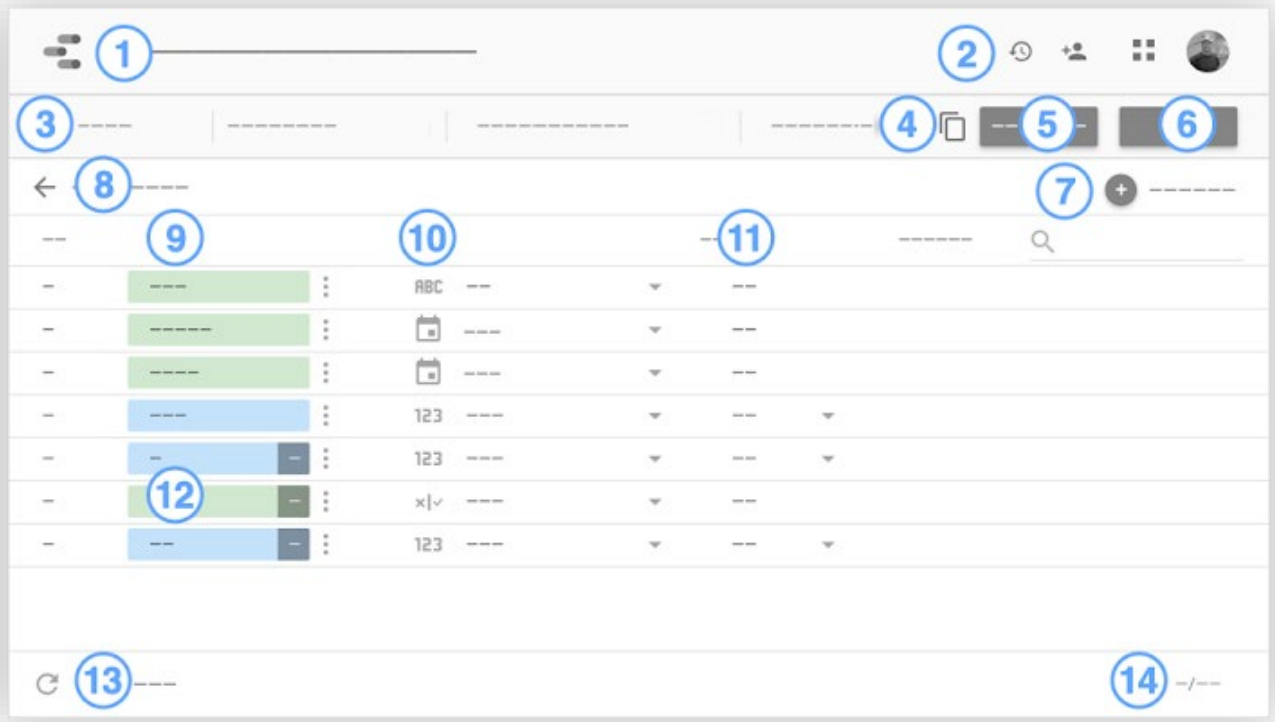
1. **Logo**. Click to return to the *Reports Home* page.
2. **Menu bar**. Many menu functions can also be accessed by right-clicking a component.
3. **User Action Menu** (From left to right:)
   - ***Refresh*** *data*
   - ***Share*** *the report: Invite other people, Schedule email delivery, Get report link, Download report.*
   - ***View****: switch between* ***edit and view mode****.*
   - *Select* ***Google Marketing Platform products***
   - ***Help***
   - *Manage your* ***Google account***
4. Manage **report pages**.
5. Selection mode | **Undo** | **Redo**.
6. **Add a chart** to your report.
7. **Add** interactive **viewer controls**.
8. **Add text, images, lines**, and **shapes**.
9. **Add data** to the report.
10. Open the **Theme and layout** panel.
11. **Change** the selected chart's **visualization type**.
12. **Select** a **chart to display** the properties panel for that chart. (To configure the chart's appearance, click *STYLE*.)
13. **Configure** the **chart's data properties**.
14. **Available fields**. Drag and drop dimensions and metrics into the data properties panel.



Find your way around the data source editor

To edit a data source used by your report:

1. Edit your report.
2. In the menus, select Resource > Manage added data sources.
3. Locate the data source in the list, then to the right, click Edit.



1. **Data source name**.
2. **Version** history.
3. **Data source options**:
   - *Data credentials.*
   - *Data freshness. .*
   - *Community visualizations access.*
   - *Field editing in reports*
4. **Make a copy** of this data source.
5. **Create report**.
6. **Explore**.
7. **Add a field**.
8. **Edit connection**.
9. **Fields**.
10. **Field type**.
11. **Aggregation**.
12. **Calculated fields**.
13. **Refresh fields**
14. **Field count**.

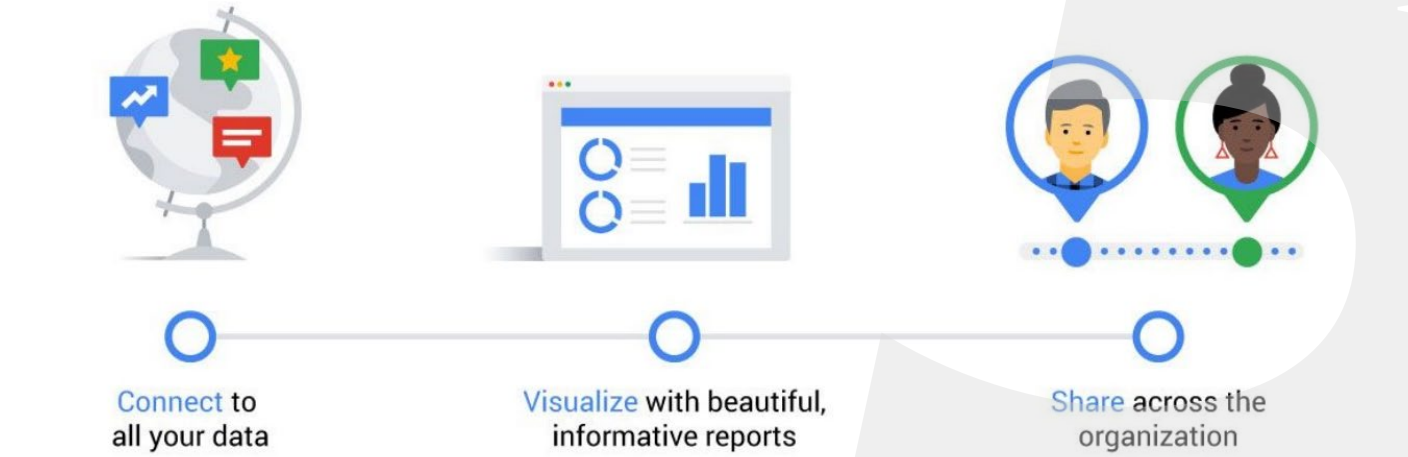
What it does?

The basic workflow of google data studio:

Connect – Acquire, prepare, transform and integrate the data.

Visualize – Collaborate to **create visual stories** with the data.

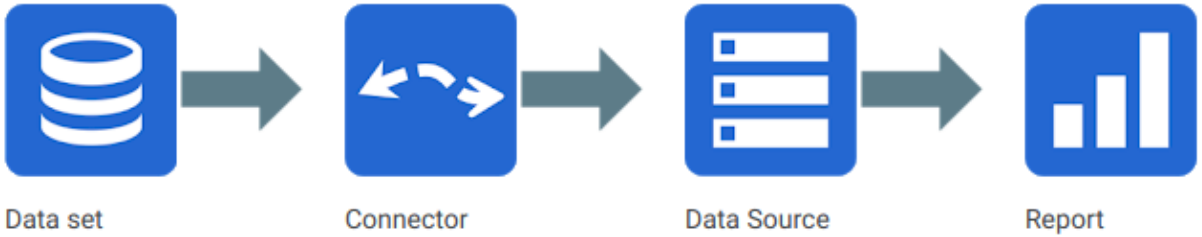
Share - Can **share reports**, insights with stakeholder, decision maker and colleagues across organization.



Concepts: **Data set, Connector & Data Source**

GDS has **4 important concepts**: Dataset, Connector, Data Source, Report.
Google Data Studio concepts and how they linked together

Data set is the "physical" layer underlying everything (and store the data), while **Data Source** is the "logical" layer with added properties and functionalities. A **Connector** is the "pipe" that connects these two layers.



A **Data set** can be much more than just a table or an Excel file.

Connector: GDS does not import your data - it uses a **connector** to get access to your actual underlying data.

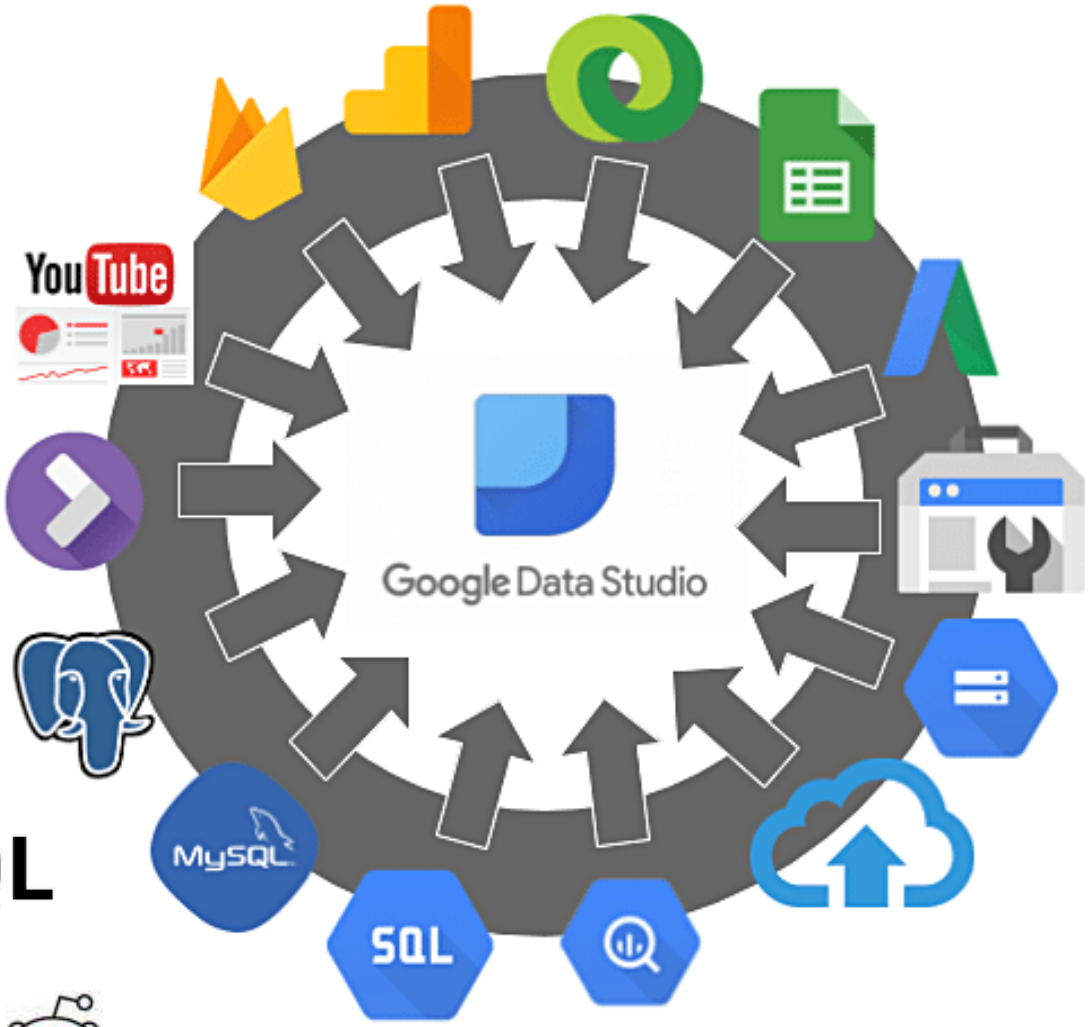
Data Source is created on top of the Data set with added functionalities such as **Shareable or Configurable.**

A **report** is the final visual presentation of data from different data sources.

What it does?

Connect to your data

- **Connecters** are required to get started.
- **Connectors** are the **link between Raw data** and **studio datasets**.
- Google Data Studio can connect and collect the data from different sources and give you one consolidated report and dashboard on single platform.



JotForm tumblr. LinkedIn SQL twitter Advertising bing Ads facebook reddit

What it does?

Easy to Connect to Data

Powerful Connectors

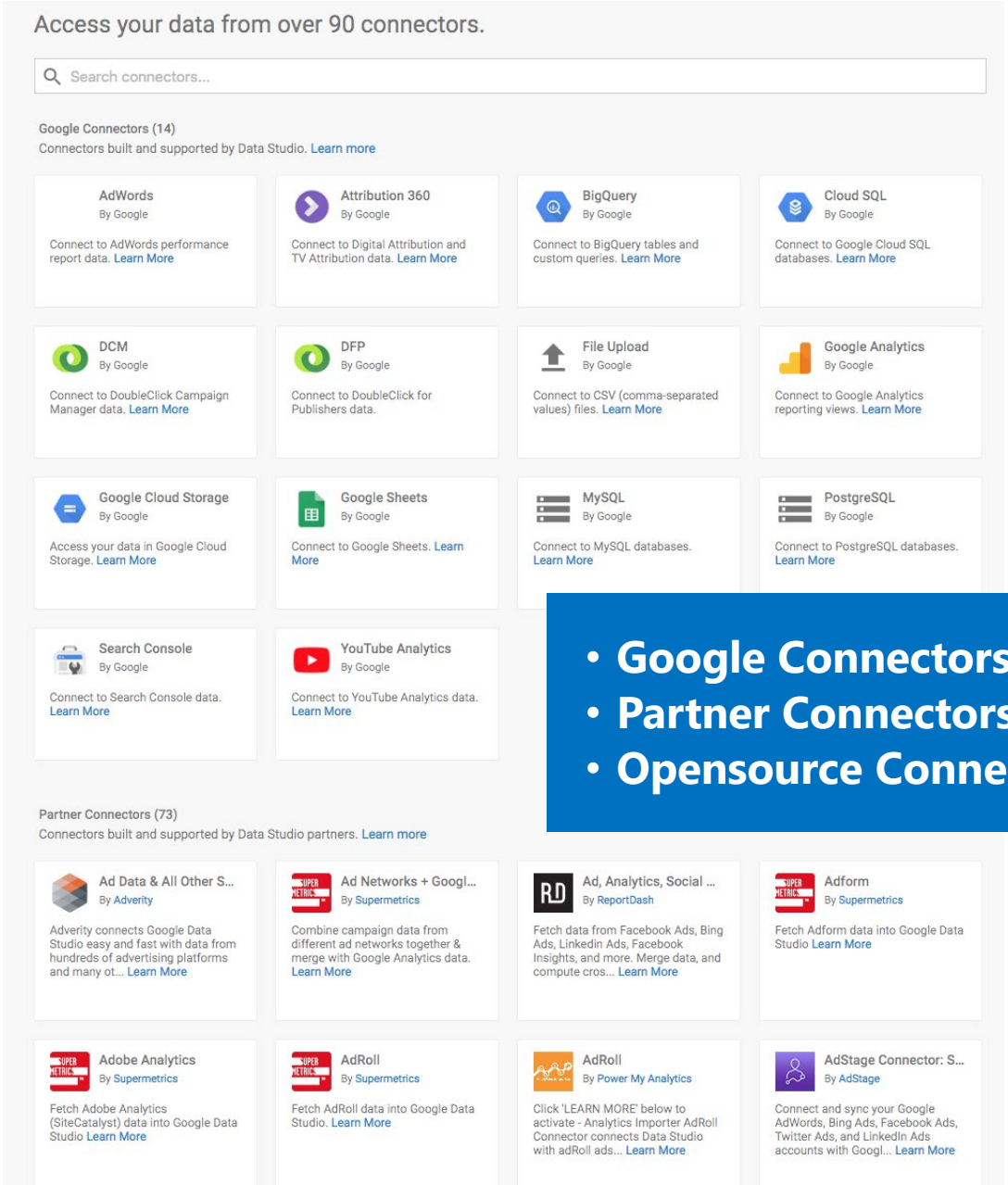
- Google Analytics
- Google Sheets
- BigQuery
- AdWords
- YouTube Analytics
- Search Console

On Premise Data

- MySQL
- PostgreSQL

Import CSV Data

- 2GB Free



- **Google Connectors (18)**
- **Partner Connectors (275)**
- **Opensource Connector (3)**

What it does?

Start with your data set

- To use Data Studio, you'll start with a ***data set*** -- a system outside of Data Studio that contains the information on which you want to report. Your data set will contain two types of data:
  - **Dimensions** are data categories. Values may include names, descriptions, date information, or other characteristics of a category.
  - **Metrics** measure the things contained in dimensions. Metrics in your Data Studio reports are aggregated, for example, as sums, counts, or ratios.
- Data sets are typically organized into rows and columns, where each row in the table contains one record, or instance, of the data.
- You may have a data set like this in a spreadsheet or it may be generated by a platform you use, such as Google Analytics, MySQL, Google Ads, or any other platform available through partner connectors.

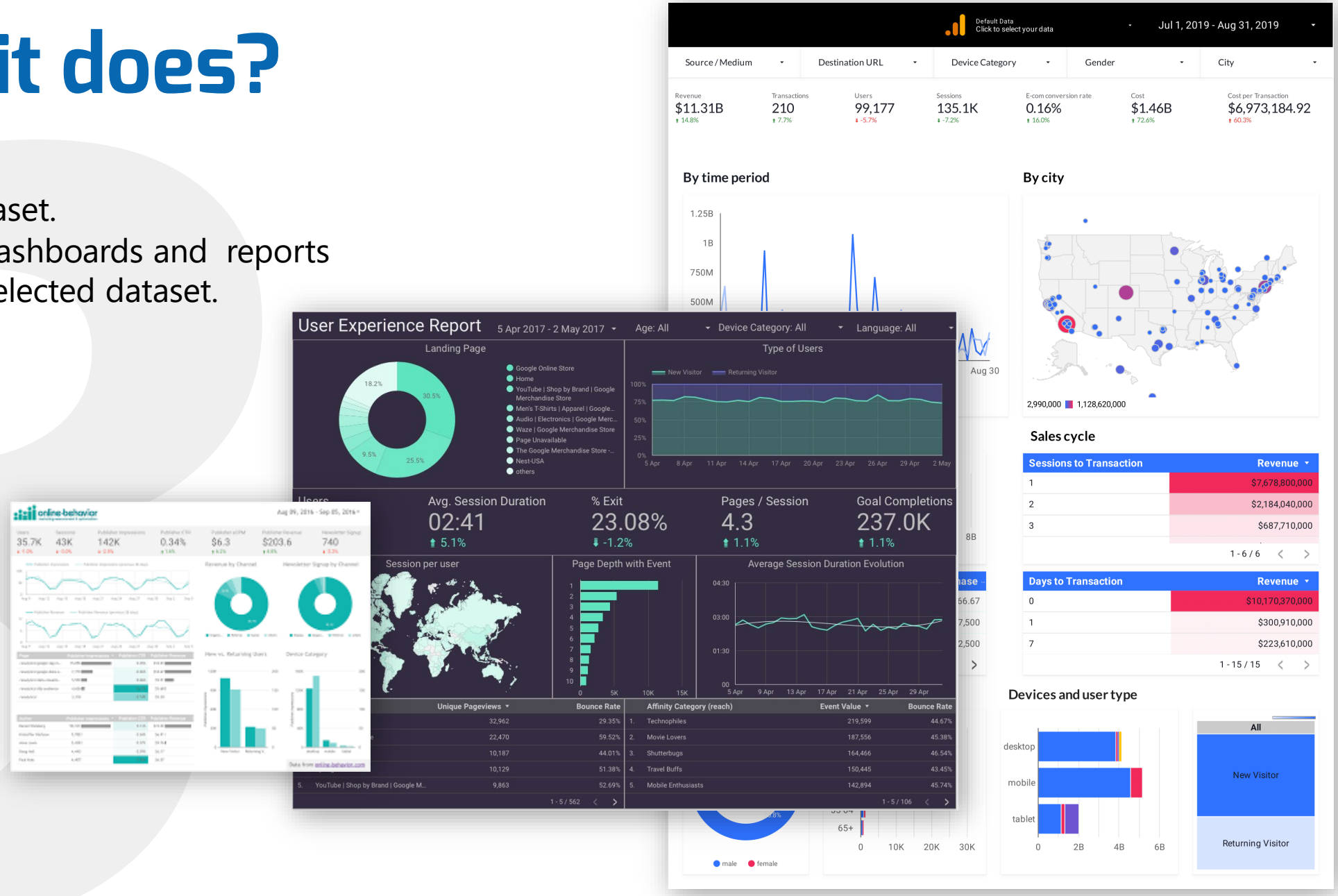
Row	Item Name	Color	Quantity Sold
1	Hat	Black	20
2	Hat	White	25
3	Coat	Red	35
4	Coat	Blue	10
5	Scarf	Purple	15

In this example, *Item Name* and *Color* are both dimensions, because they describe or categorize the products for sale. Dimensions can be any other kind of data, including unaggregated numbers, dates, text, and boolean (true/false) values. *Quantity Sold* is a metric; it is a count of that row's items sold.

What it does?

Visualize

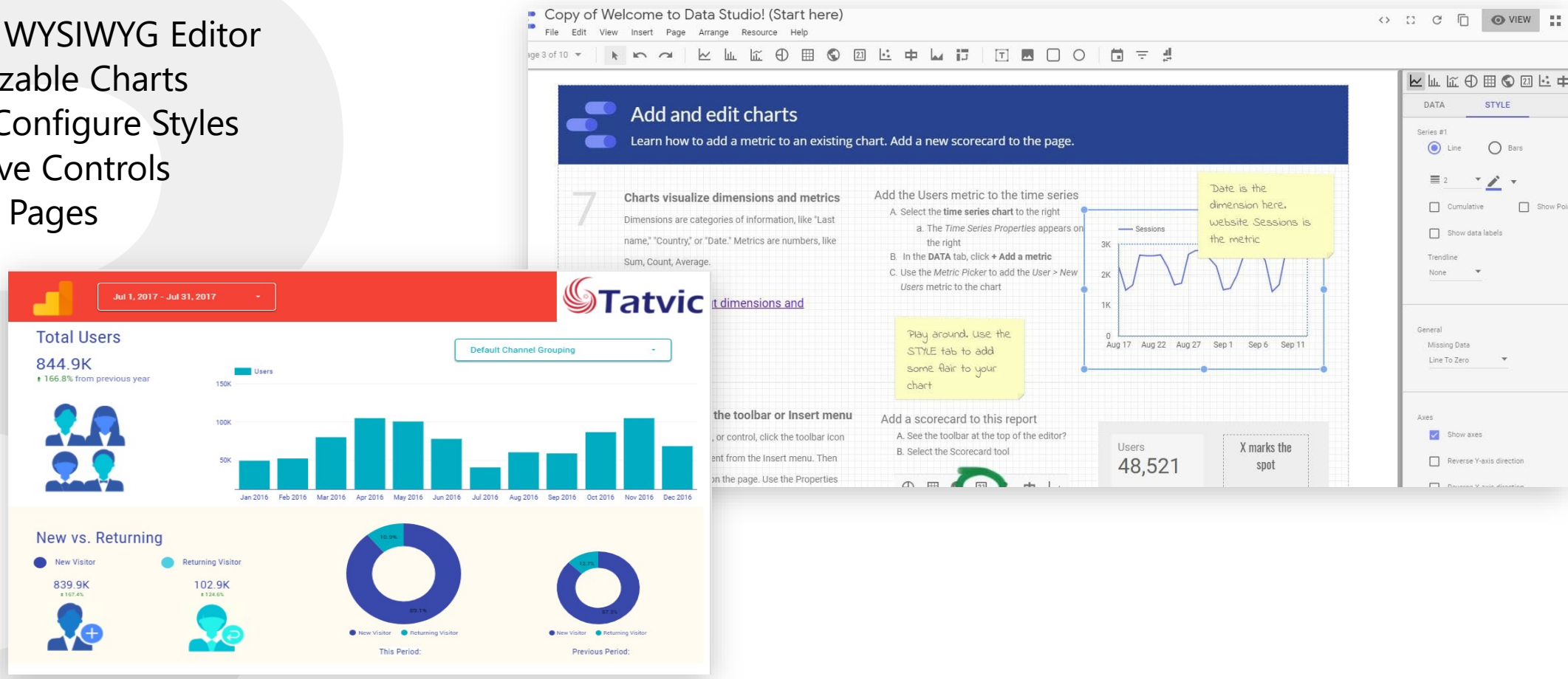
- Choose dataset.
- Generate dashboards and reports based on selected dataset.



What it does?

Visualize - Build Beautiful Reports

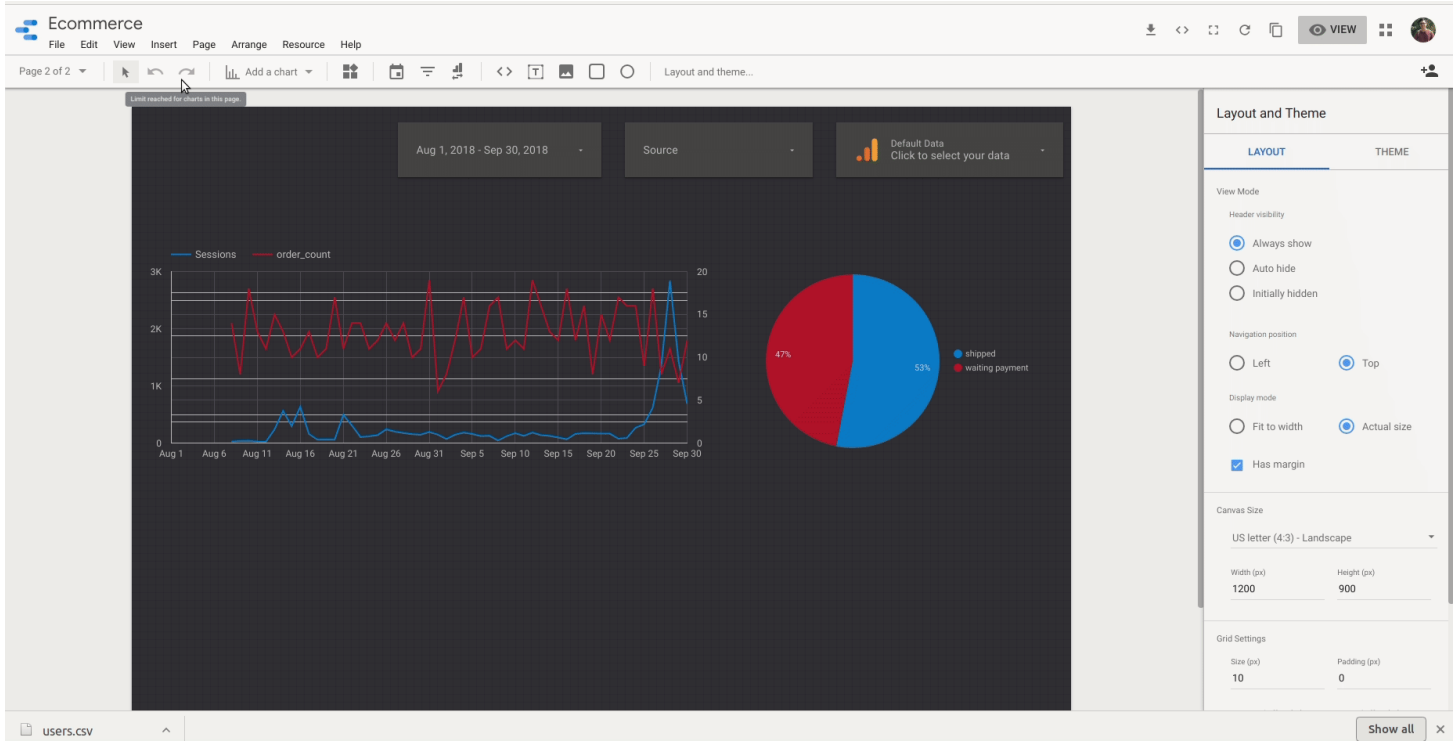
- Modern WYSIWYG Editor
- Customizable Charts
- Easy to Configure Styles
- Interactive Controls
- Multiple Pages



Visualize your data

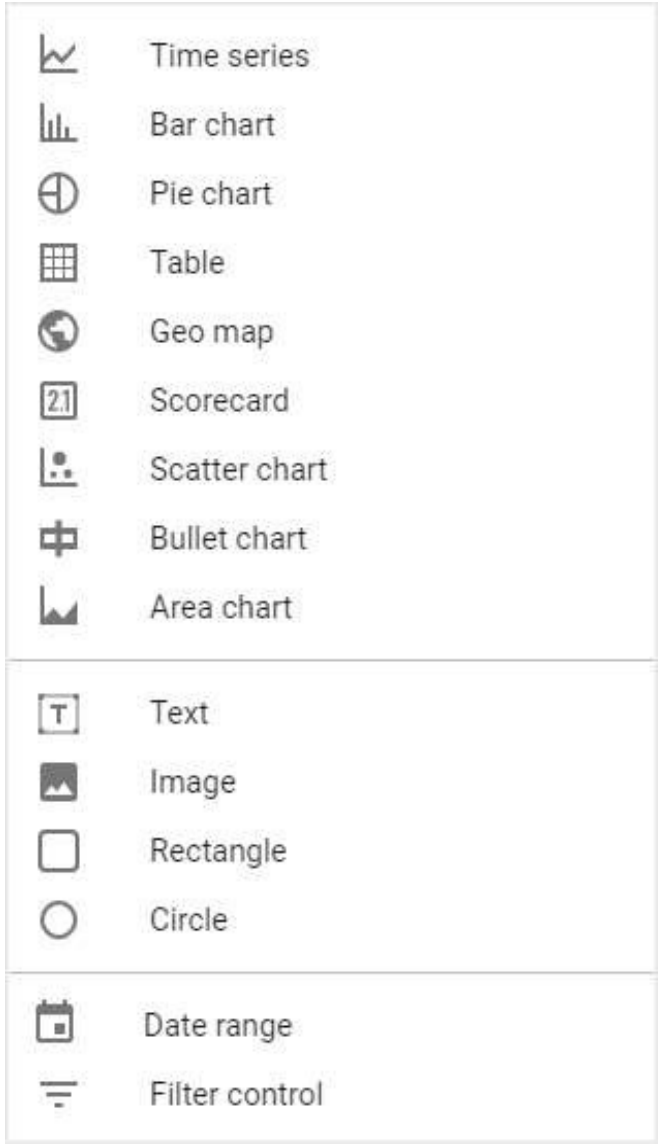
Use the drag and drop report editor to:

- Tell your data story with **charts**, including **line**, **bar**, and **pie charts**, **geo maps**, **area** and **bubble graphs**, **paginated data tables**, **pivot tables**, and more.
- Make your reports interactive with viewer **filters** and **date range controls**. The data control turns any report into a **flexible template** report that anyone can use to see their own data.
- Include **links** and **clickable images** to create product **catalogs**, **video libraries**, and other **hyperlinked content**.
- **Annotate** and **brand** your reports with text and images.
- Apply **styles** and **color themes** that make your data stories works of data visualization art.



Tools

- Drag and move items around
- the report.
- draw shapes and text. import images.
- customize colors and settings.

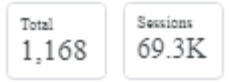


Tools

An overview of all Google Data Studio Chart types in 2019

Tables

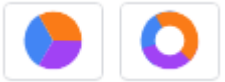
Scorecards



Time Series

Bar charts

Pie charts



Geo maps

Line charts

Area charts

Scatter charts

Pivot tables

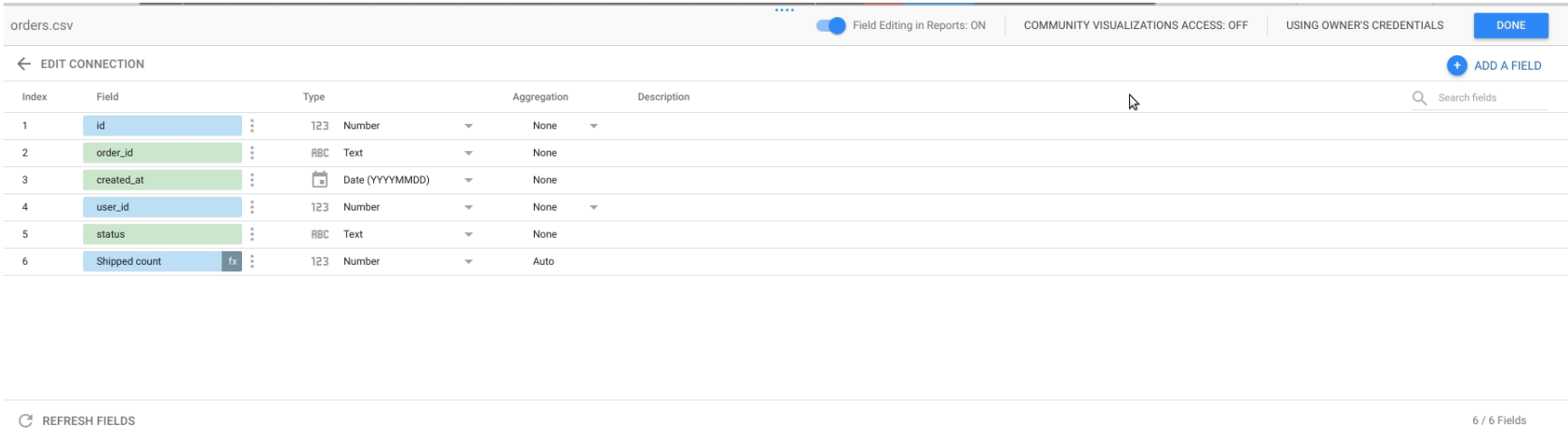
Bullet charts

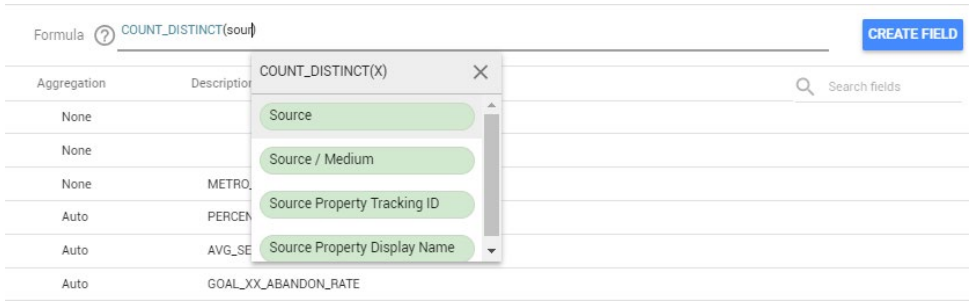
Treemaps

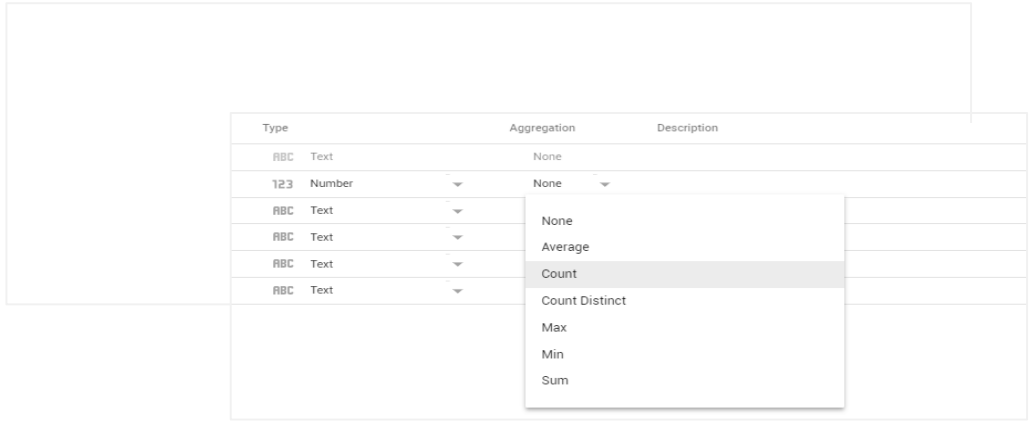
What it does?

Visualize - Custom Data Transformations

- Aggregations
- Groupings
- Ad Hoc Calculations







What it does?

Type of Data

Data Type		Meaning What?	Example
Categorical (Dimensions)	Nominal	The ordering of values has no intrinsic meaning	Event Category, User Type, City
	Ordinal	Ordered, "Every value has a place"	Count of Session, Month of the Year
	Time Series	Dates and Times	Date, Hour
Quantitative (Metrics)	Additive	Adding values makes sense, zero is meaningful	Sessions, Transactions
	Non-additive	Only difference in values makes sense	Avg. Session Duration, Bounce Rate

What it does?

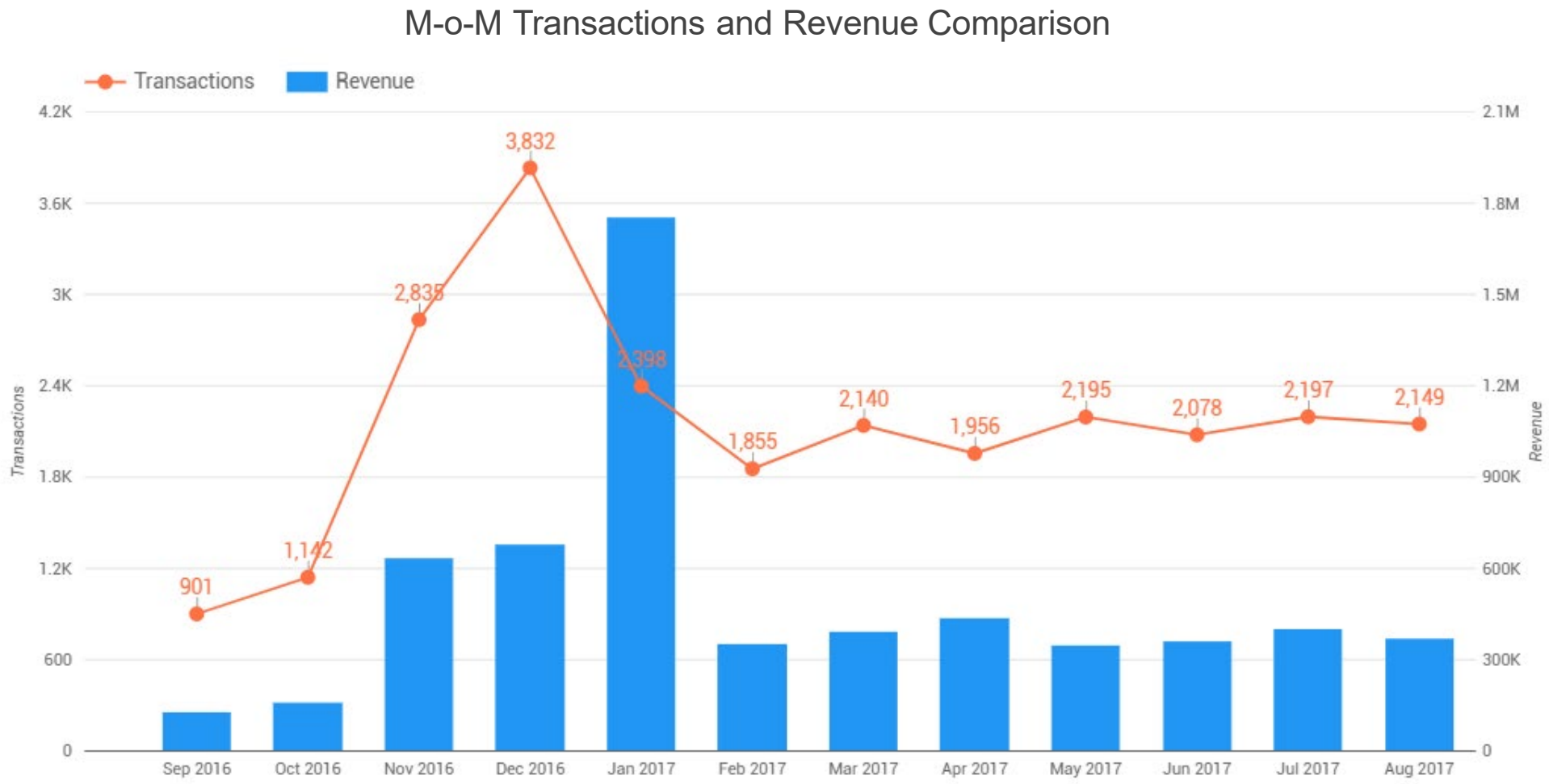
Tabulate Relationships Between Metrics

Source/Medium wise Metrics Relationship

Source / Medium	Sessions	% Δ	Transactions	% Δ	Revenue	% Δ	Ecommerce Conversion Rate	% Δ
google / organic	38,455	5.0%	622	-0.3%	$103,088.70	7.9%	1.62%	-5.0%
(direct) / (none)	11,358	5.3%	136	22.5%	$23,802.09	-61.7%	1.2%	16.3%
youtube.com / referral	10,418	76.4%	1	0.0%	$1.99	-43.0%	0.01%	-43.3%
google / cpc	4,333	44.1%	36	-14.3%	$2,632.98	-49.6%	0.83%	-40.5%
m.facebook.com / referral	3,870	1,575.3%	1	-	$6.99	-	0.03%	-

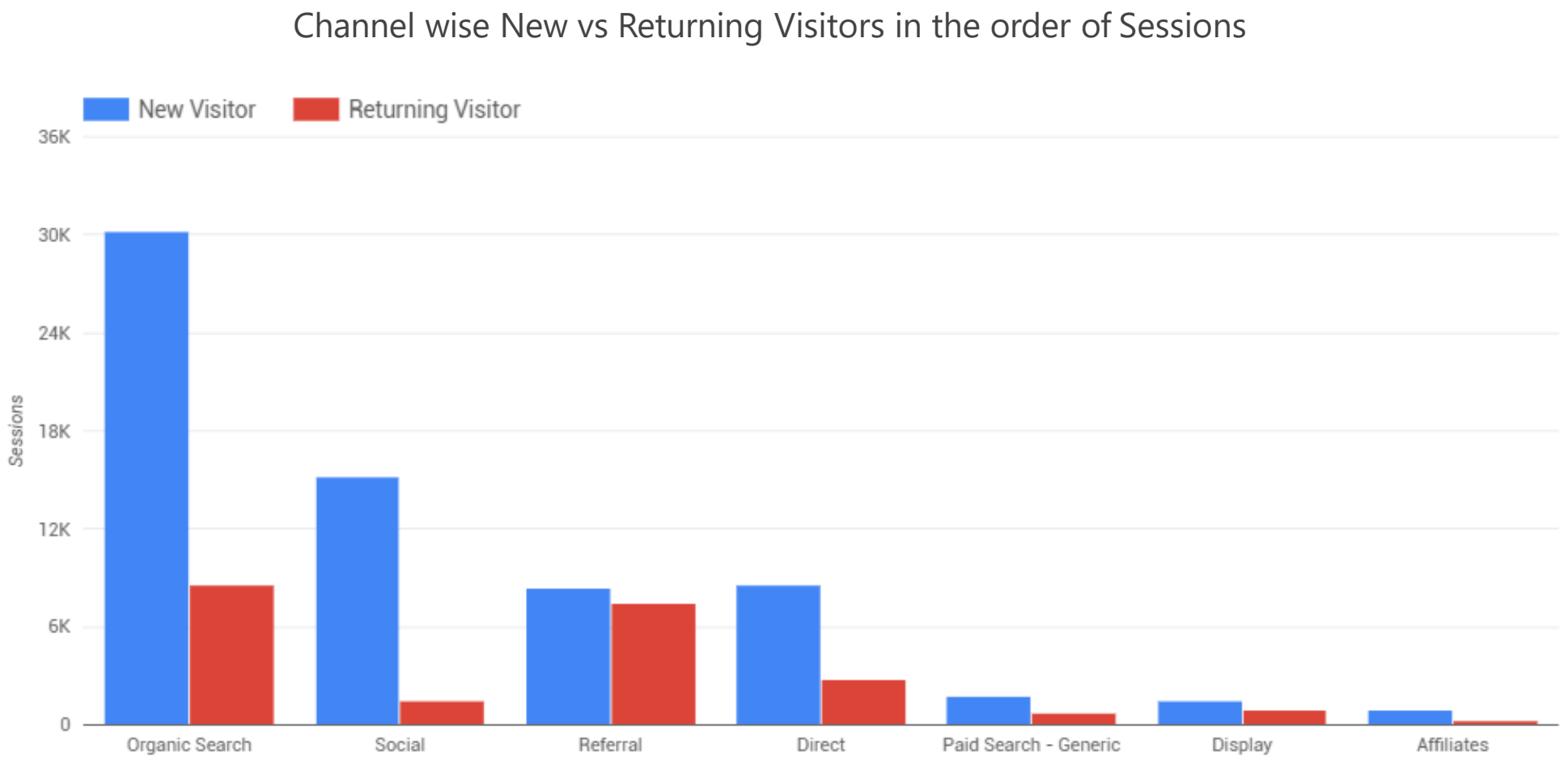
What it does?

Compare two related Metrics



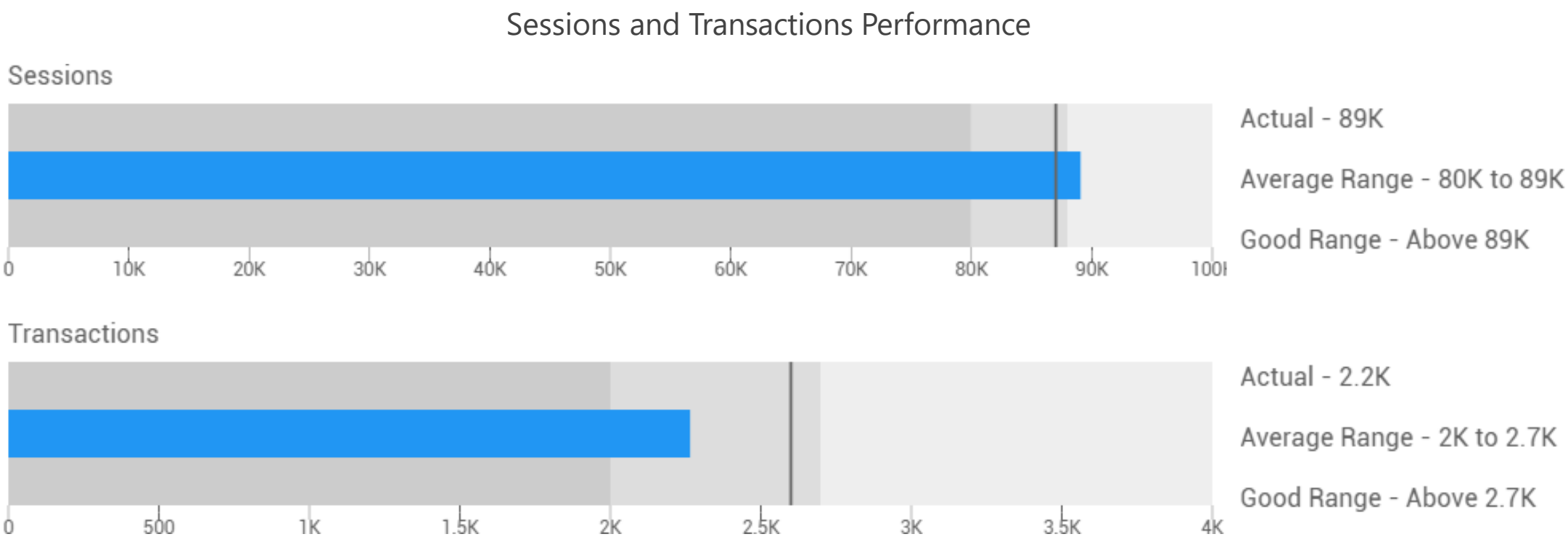
What it does?

Visualize two Dimensions over a Metric



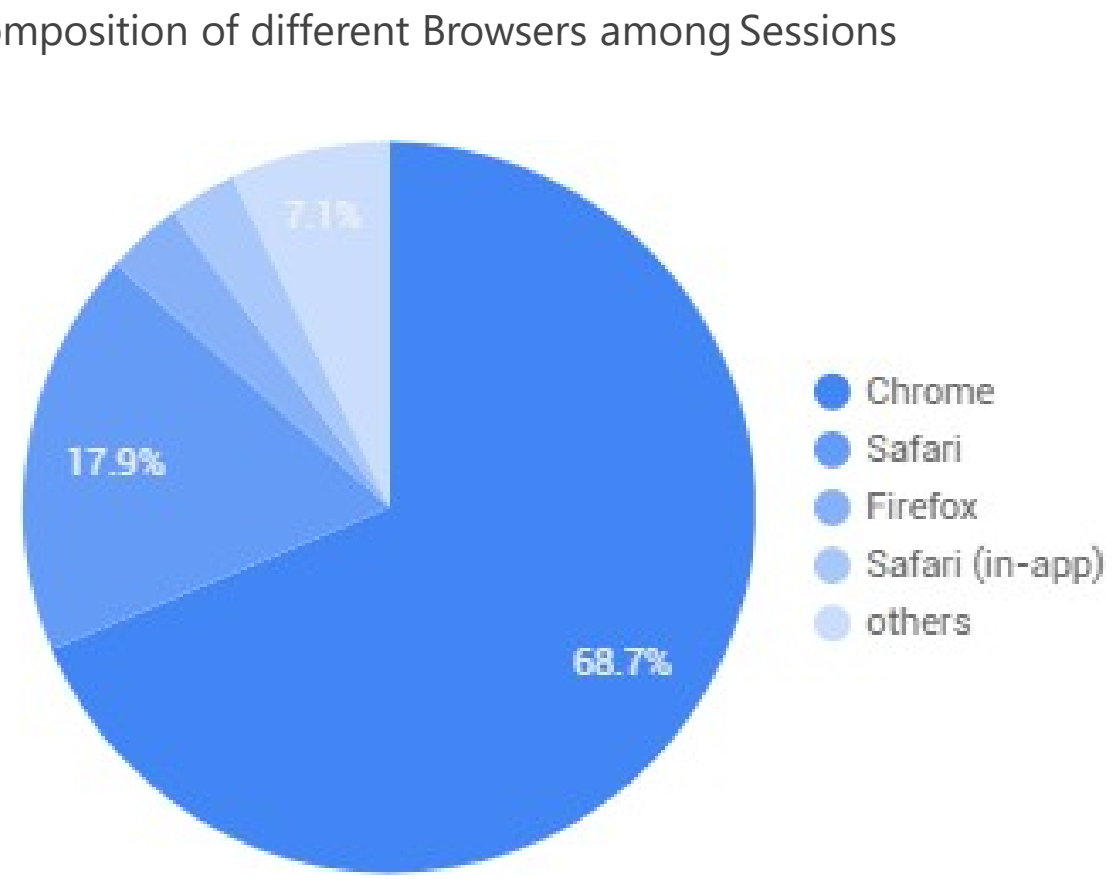
What it does?

Monitor Performance of a Metric



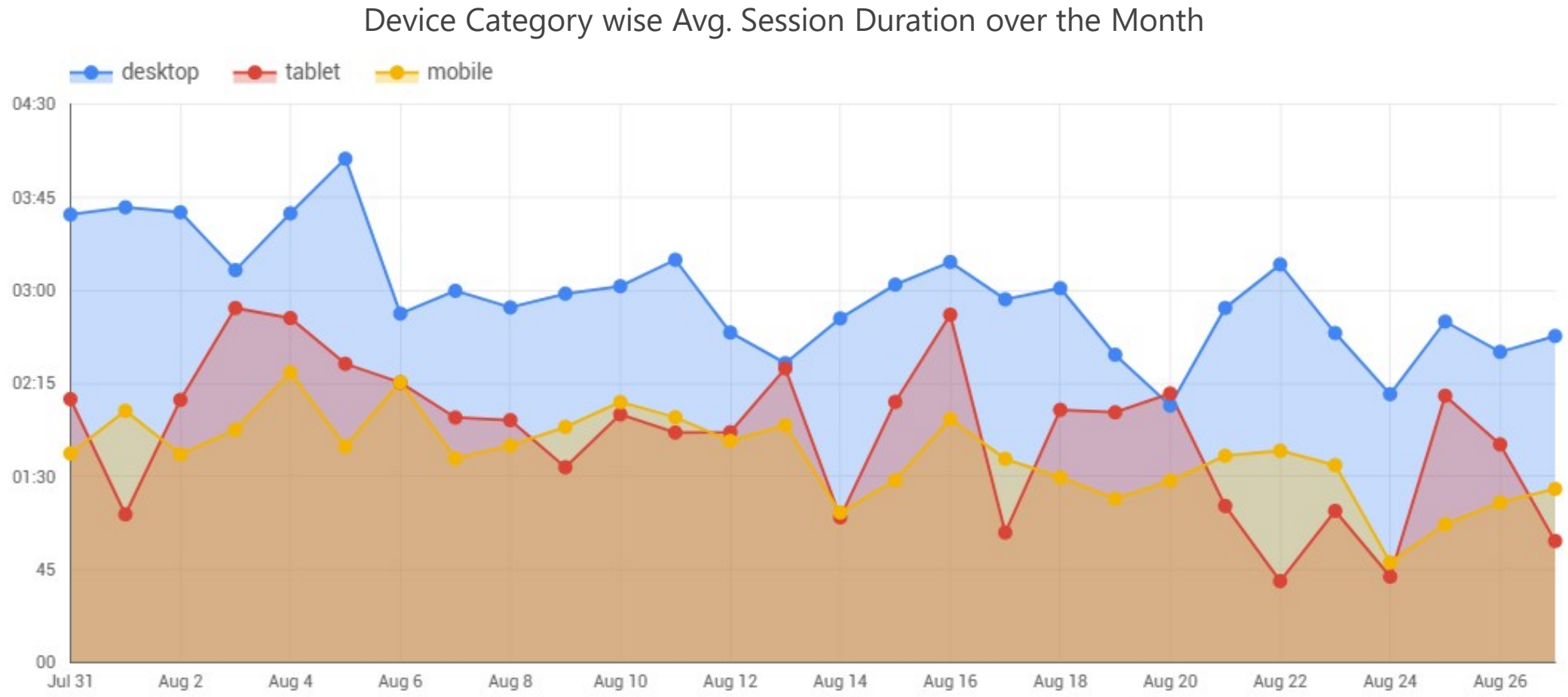
What it does?

Show Dimension Composition w.r.t. a Metric



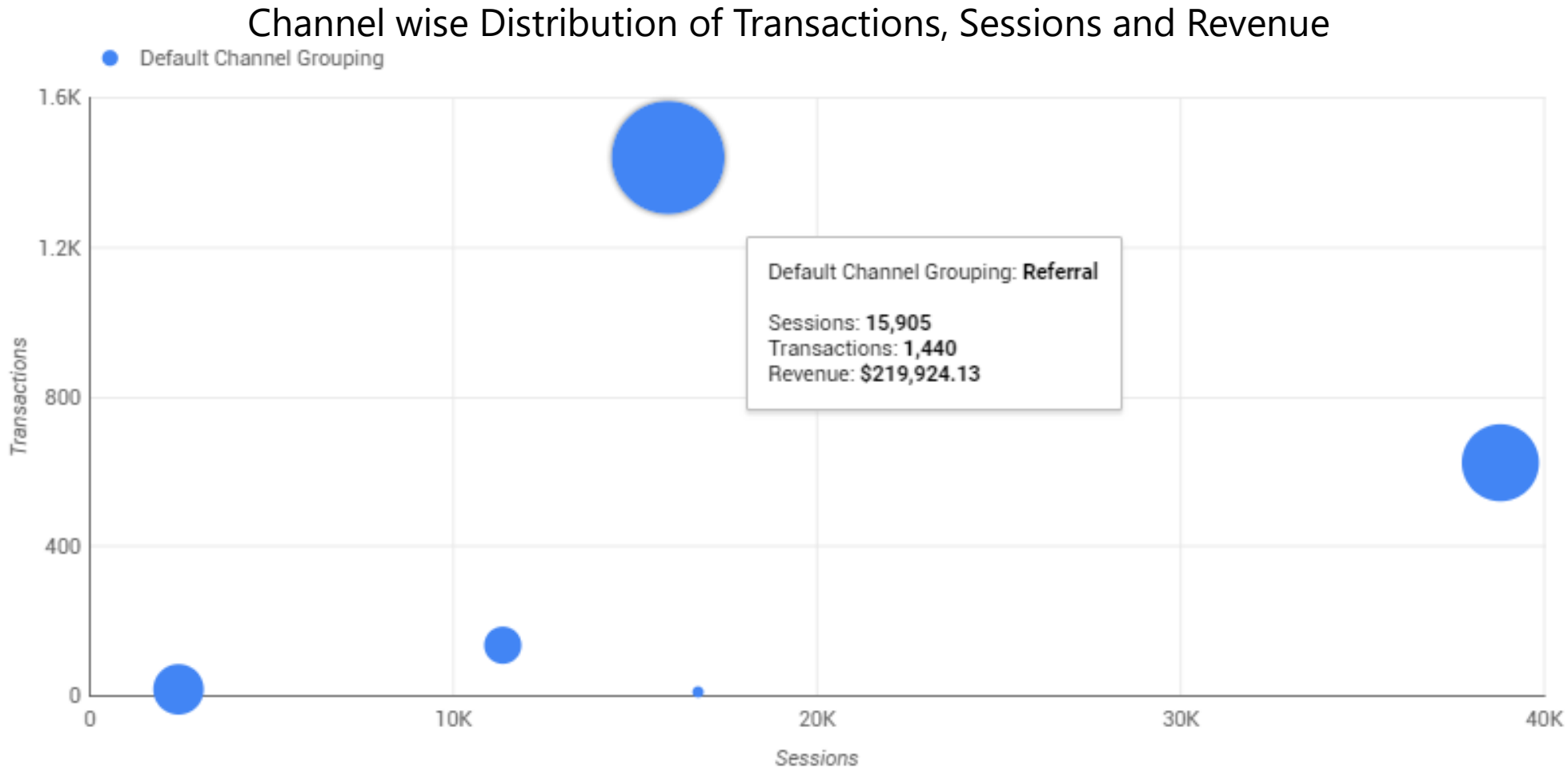
What it does?

Observe Metric Trend w.r.t. a Dimension



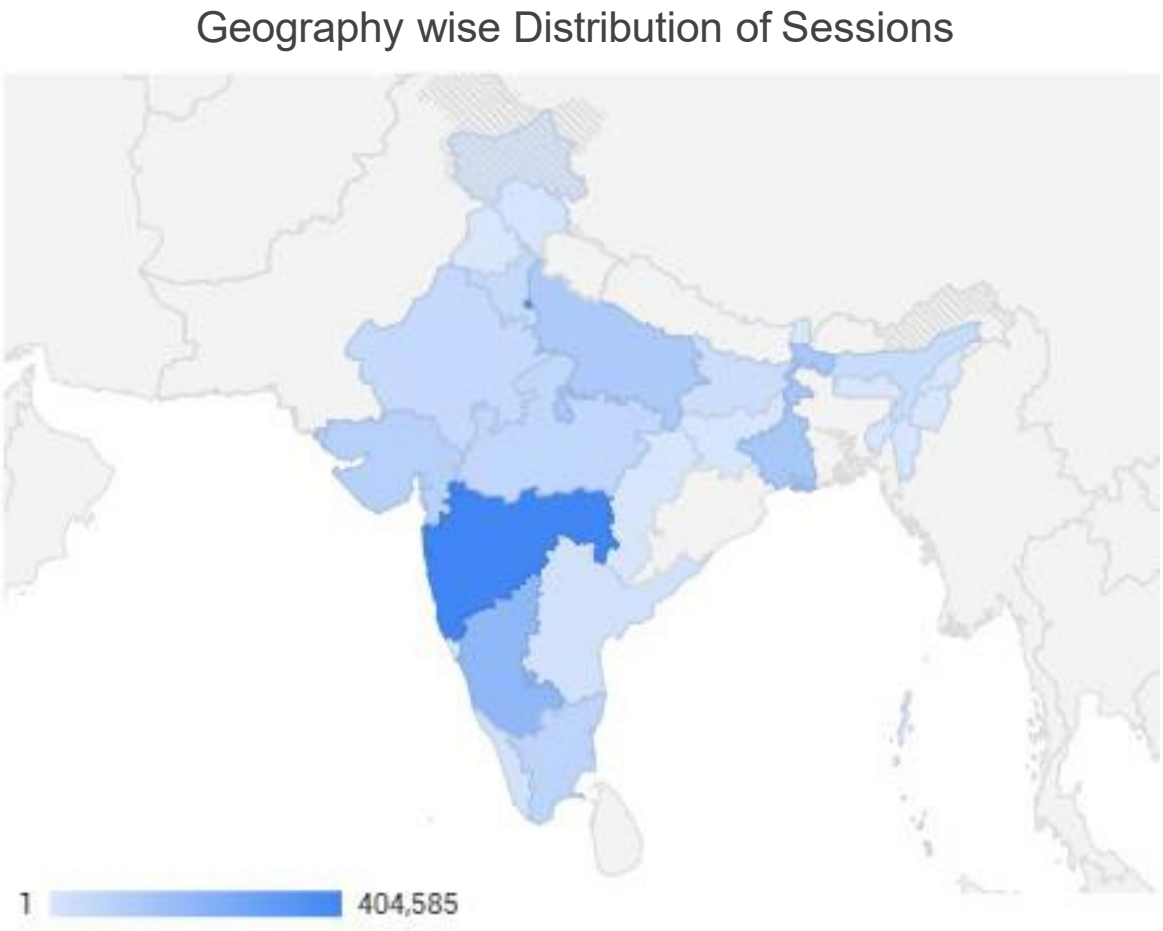
What it does?

Analyze Metrics Scatter Plot for a Dimension



What it does?

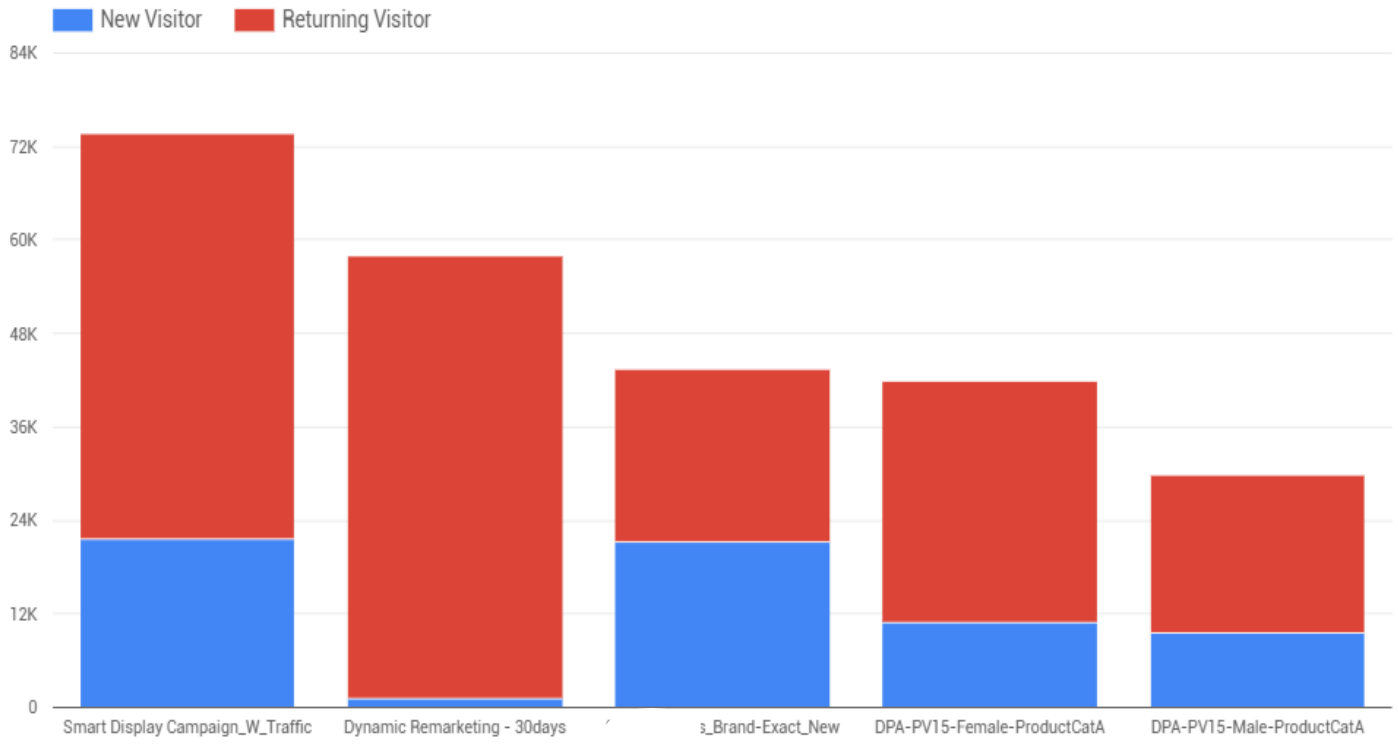
View Geo-Distribution of a Metric



What it does?

Breakdown Dimension w.r.t. another Dimension Camnpaign

Breakdown w.r.t. Sessions from User Type



Tools

Reports Best

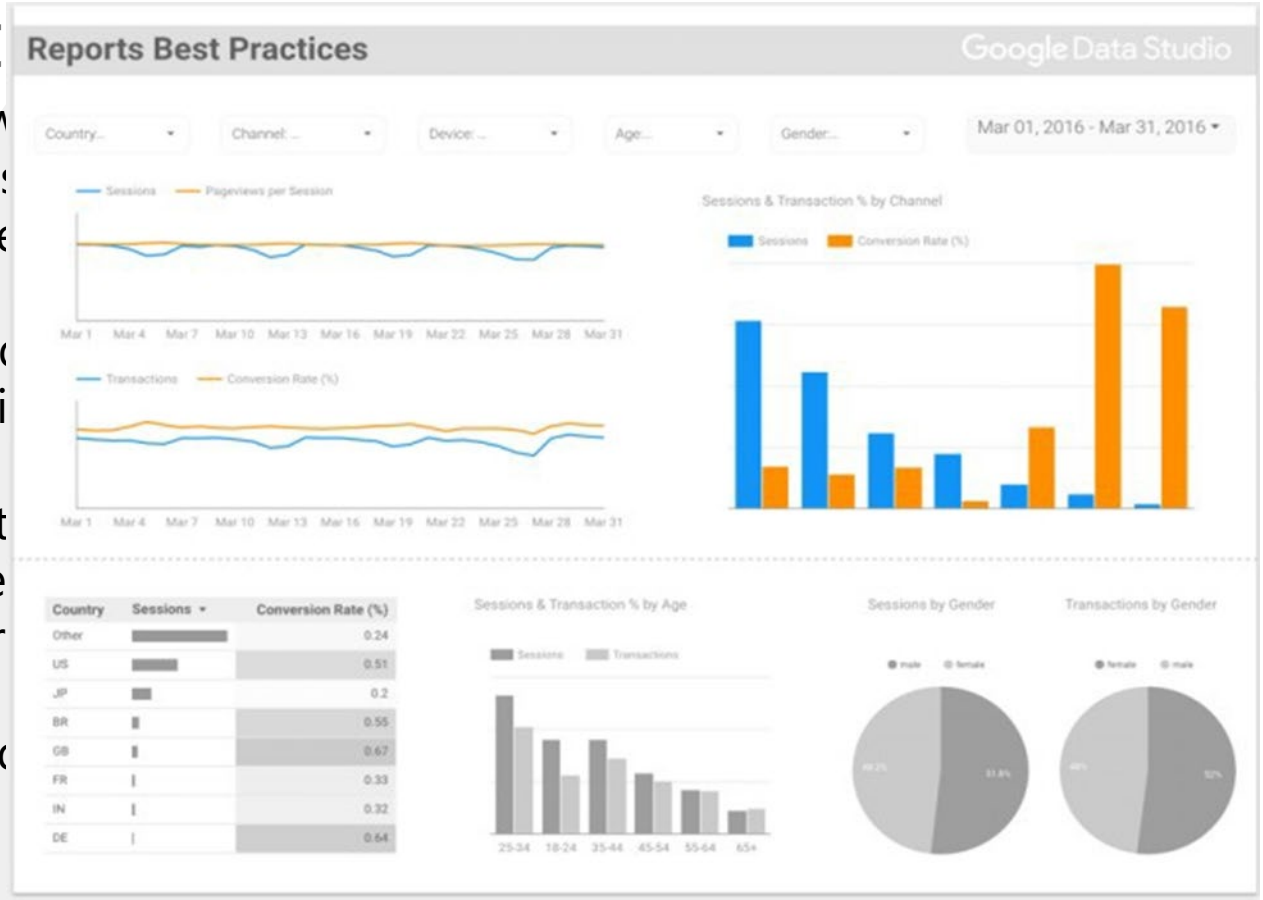
Filter controls give pow
controls will drive user
rich analysis experience

Headers and page divi
organization and consi

Chart diversity makes t
engaging. Having diffe
the analysis more inter

Color styling helps gui
overuse color!

The Report purpose informs the design. The purpose of the visualizations is incredibly important during the conception and creation of your Reports.



- Time series
- Bar chart
- Pie chart
- Table
- Geo map
- Scorecard
- Scatter chart
- Bullet chart
- Area chart

- Text
- Image
- Rectangle
- Circle

- Date range
- Filter control

Sample Infographic

https://datastudio.google.com/s/rfrsXbJ8_E0

What it does?

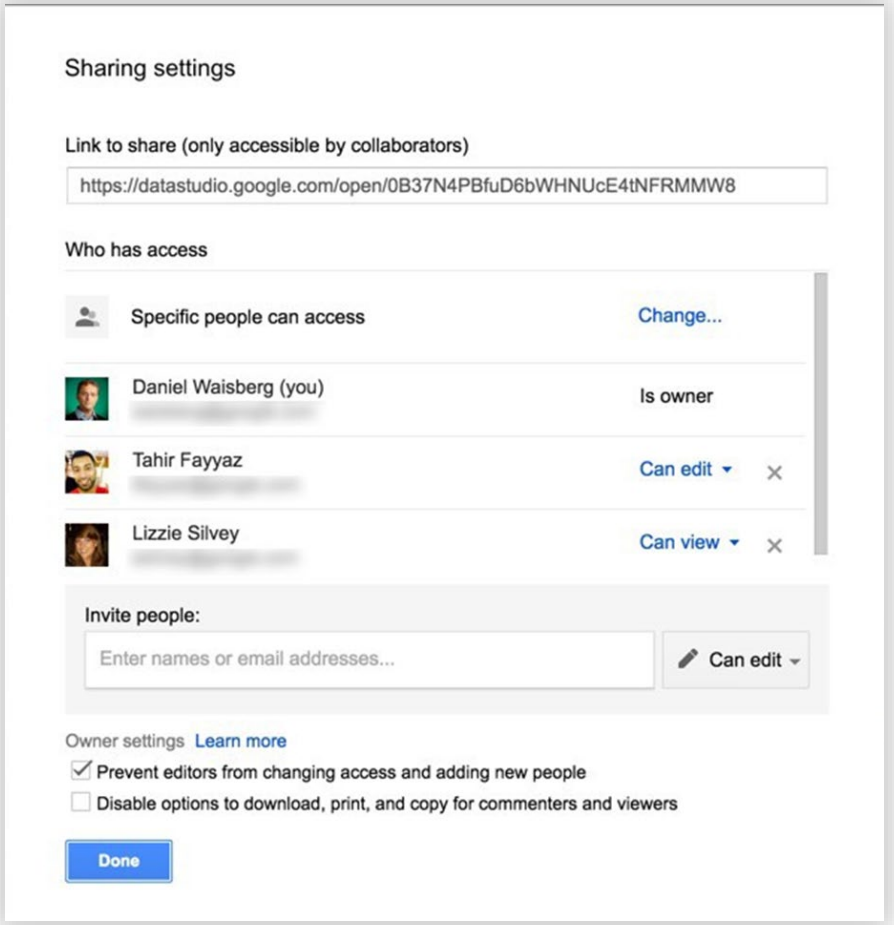
Share

- once report generated,
- just share it like normal Google dri documents.

Share and collaborate

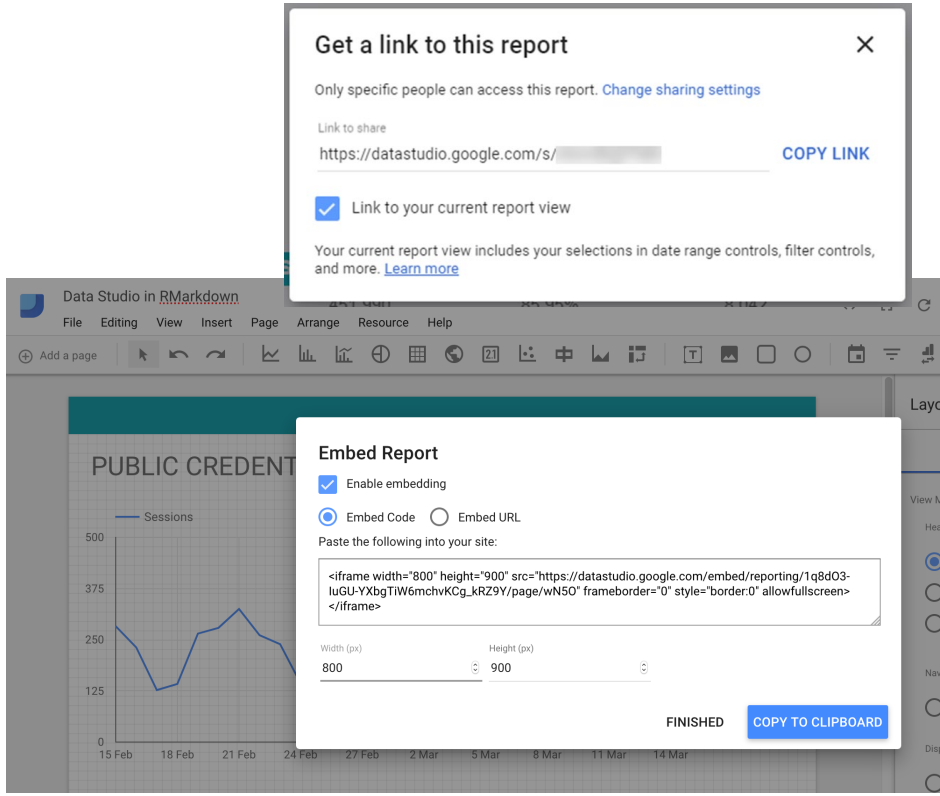
Sharing the report:

- Data Studio uses the **Google Drive sharing model**, which you are hopefully acquainted with.
- It is important to notice that when you share a Report or Data Source with a person, the access will be given unrelated to whether the person has access to the data in Google Analytics, Sheets, BigQuery, etc.
- This means that it is extremely **important to make** sure that the **data can be shared with the person**.



Share and collaborate

- It's easy to share your insights with individuals, teams, or the world. Invite others to view or edit your reports, or send them **links** in **scheduled emails**.
- To tell your data stories as broadly as possible, you can **embed** your reports in other pages, such as Google Sites, blog posts, marketing articles, and annual reports.
- When you share a Data Studio file with another editor, you can **work it together in real time as a team**.



Share securely

- to keep your Data Studio reports **private**, **share** them with a **limited group of people**, or **share them with the world!**
- There are a few access settings to keep in mind as you use Data Studio:
  - "***Permissions***" *determine who can view or edit your Data Studio files -- that is, your data sources and reports.*
  - "***Credentials***" *refers to who can see the data within those reports.*

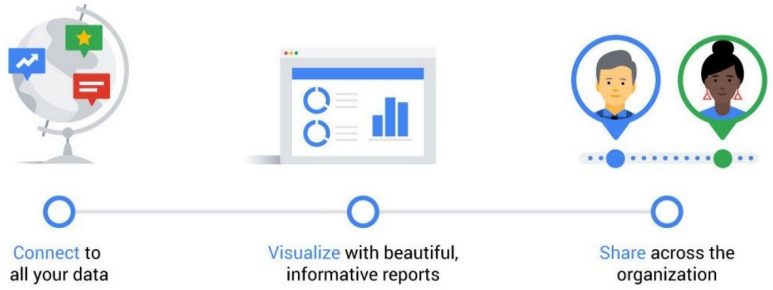
Share and collaborate

Review access controls

- To determine whether *others* can view report data, the data source credentials setting has two options: Owner's and Viewer's:
  - ***Owner's Credentials*** *uses the credentials of the data source owner to allow others to view data in reports.*
  - ***Viewer's Credentials*** *requires that viewers have their own access to the data in order to be able to see data in reports.*

Data Storytelling & Visualization Competition

Objective



To optimize **Google Data Studio** application usage among UiTM Staff and Students.

Powered by Google Data Studio

To encourage UiTM Staff and Students **discover** and **share insights** from data that can be accessed on any device, anytime and anywhere.

Eligibility

1

The contest is **open to all UiTM staff and students**.

2

All entries should be submitted on **individual basis**.

3

Only **one entry** is allowed **per participant**.

Data Storytelling & Visualization Content

DSTV Content designs should:

1

produce DSTV **related to day to day work/study** that **benefit to UiTM Staff, Student** or **Community.**

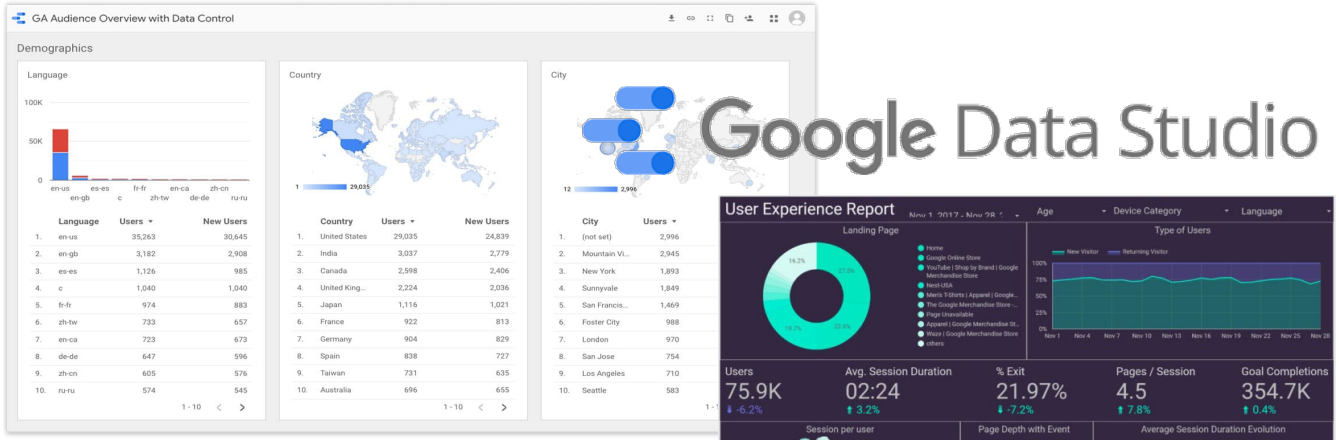
2

contained scope of data that focus **on the specific topic** or **theme per entry**

3

Each entry should produce **one (1) report** with **not more than 10 pages**

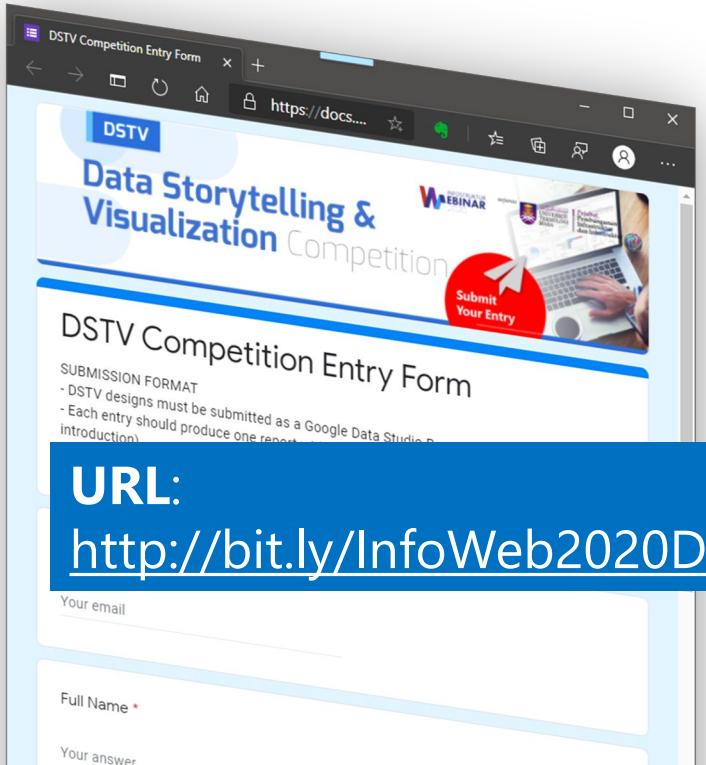
Entry Submission



DSTV designs must be submitted as a **Google Data Studio Report**.

Entries must be submitted to organizer by **Monday, November 30, 2020** not later than 12.00 p.m.

All DSTV entries must be submitted through **competition entry form** provided by organizer.



URL: http://bit.ly/InfoWeb2020DSTV

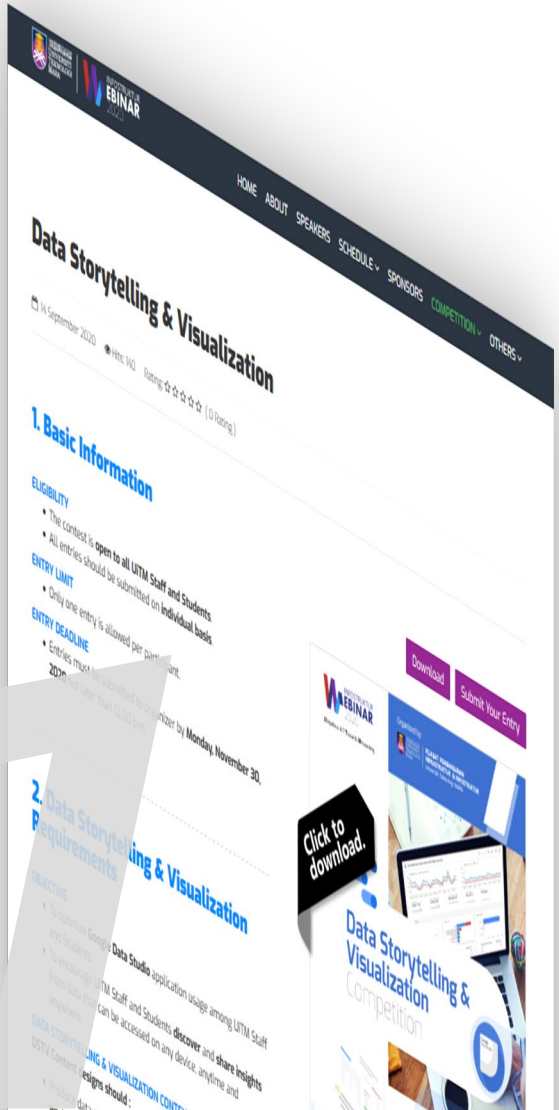
Competition Winner

1. **Three winners** for each category (Staff and Students) will be selected.
2. Winners will be selected by the **panel judges of Infostruktur Webinar 2020**.
3. These judges are selected at the sole discretion of the Infostruktur Webinar 2020.
4. Winners will be announced on **December 2020** during **Infostruktur Webinar 2020 Closing Ceremony.**

More Information

1

Go to ictevent.uitm.edu.my



2

Download DSTV Competition Flyer

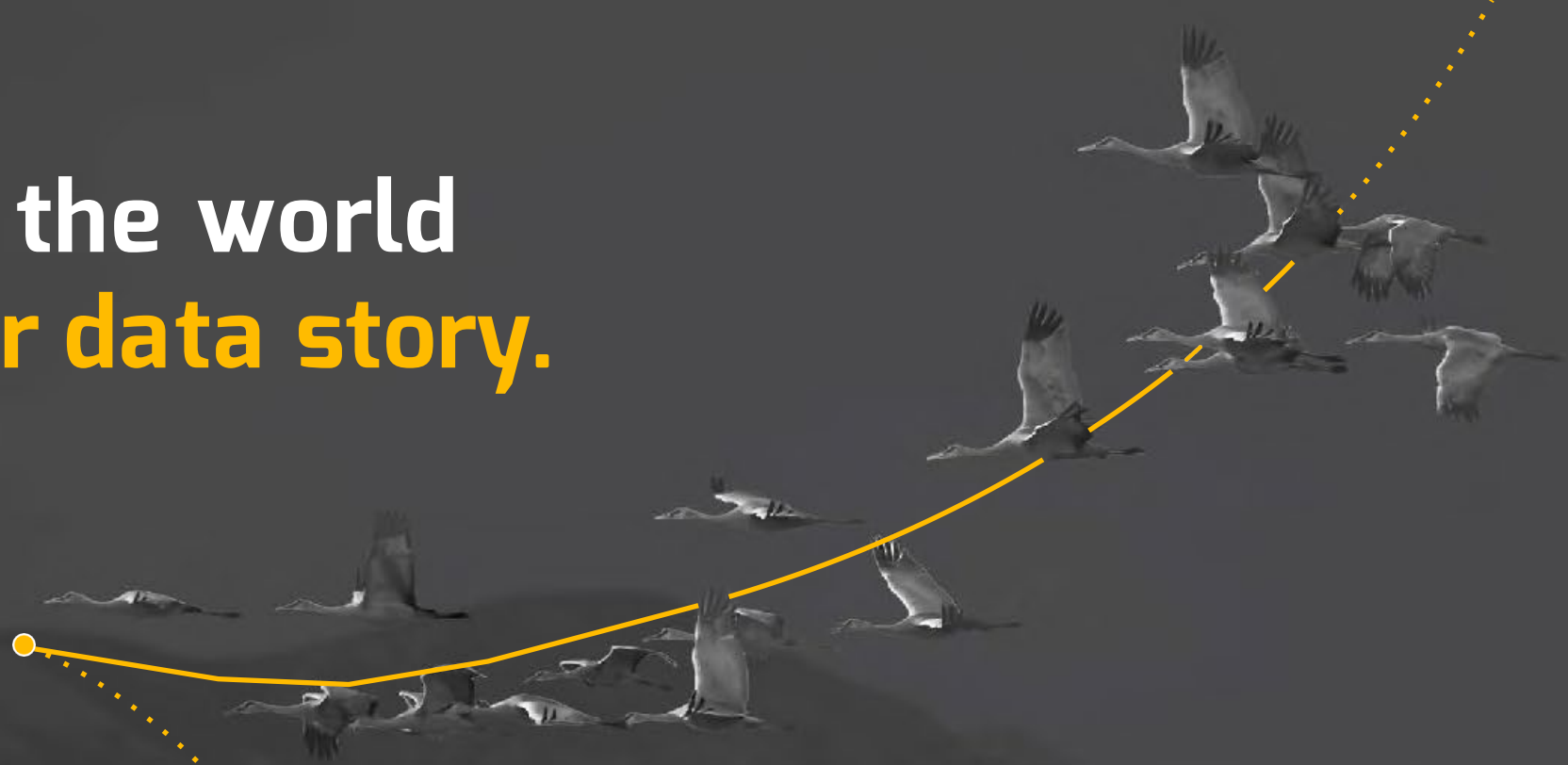
3

Contact Us

Gazairi Ghazali
03-55211341
gazairi@uitm.edu.my
Zamani Umar Husin
03-55442252
zamanium@uitm.edu.my

Conclusion

Tell the world your data story.



"The greatest value of a picture is when it forces us to notice what we never expected to see." - John Tukey

FIVE STEPS TO GIVE YOUR DATA A HAPPY ENDING

Like the sound of data storytelling but not sure how to put it into practice?
Here are **five things** you can do immediately to help create data stories that provide valuable insight to your business:

FIVE STEPS
TO GIVE YOUR DATA
A HAPPY ENDING

ONE

IDENTIFY YOUR AUDIENCE

Ask yourself:

- Who am **I reporting** to?
- How do they like to **consume information**?
- Is there just **one group** or **different audiences**?
- Where and when can **I communicate** with them?

two

ESTABLISH AN OBJECTIVE AND STORY

Ask yourself:

- What **business decisions** do my audience need to make?
- What **problems** are they trying to solve?
- What **do they already know**?
- What have **they been told before**?
- How **important** is the decision?
- Am I **recommending a decision** or providing the **facts**?

three

DECIDE WHAT DATA WILL HELP YOU TELL THAT STORY

Ask yourself:

- What **data** does the company have **available** to **investigate** the story?
- Do I need to do anything to **use** these **datasets**?
- Can I **gather new data**?
- What **analysis techniques** can I use to surface the insights?

four

DECIDE HOW TO TELL YOUR STORY

- What is the **best way** to bring my story to life for my audience?
- What **visualizations** should I use?
- What **software** do I have available?
- How often do I need to **update** the data?

five

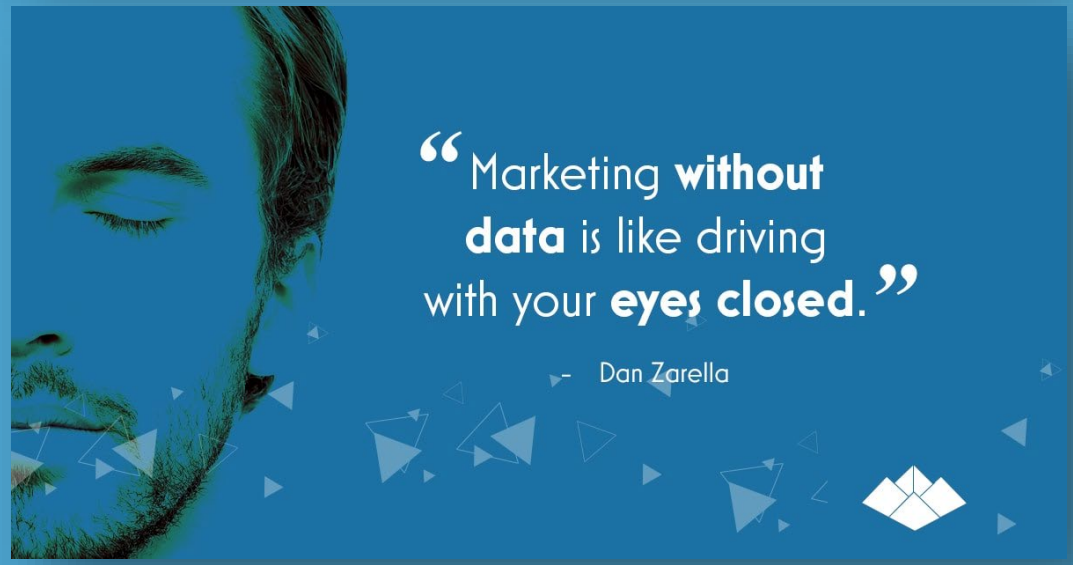
IMPROVE NEXT TIME

Ask yourself:

- Did my audience **understand** everything?
- Did I give them **sufficient information**?
- Was the **decision successful**?
- Is there anything **new to add** in the future?


Data will talk to you if you're willing to listen to it.
Jim Bergeson


Without data, you are just another person with an opinion
- W. Edwards Deming


Marketing without data is like driving with your eyes closed.
- Dan Zarella

Thank You

Contact Us

Zamani Umar Husin
03-55442252
zamanium@uitm.edu.my

Gazairi Ghazali
03-55211341
gazairi@uitm.edu.my

Please Scan **QR Code** for You
Feedback & Attendance

References

- **From Data to Decisions: The Power of Visual Storytelling**, 21st March 2019, Dr. Vinod Ramachandran-Data Scientist, theCADS
- **Data Driven Leadership & Culture,** Sharala Axryd - Founder and CEO, The Center of Applied Data Science
- **Data Storytelling: The Essential Data Science Skill Everyone Needs**, Brent Dykes, Mar 31, 2016, https://www.forbes.com/
- **Data Storytelling Handbook** - Best Practices in Data Visualization, SAP BusinessObject
- **Data Stories -** Creating compelling stories with data, February 24, 2014, Prepared by: Miriam Gilbert
- **Data Storytelling** – Bright North, UK