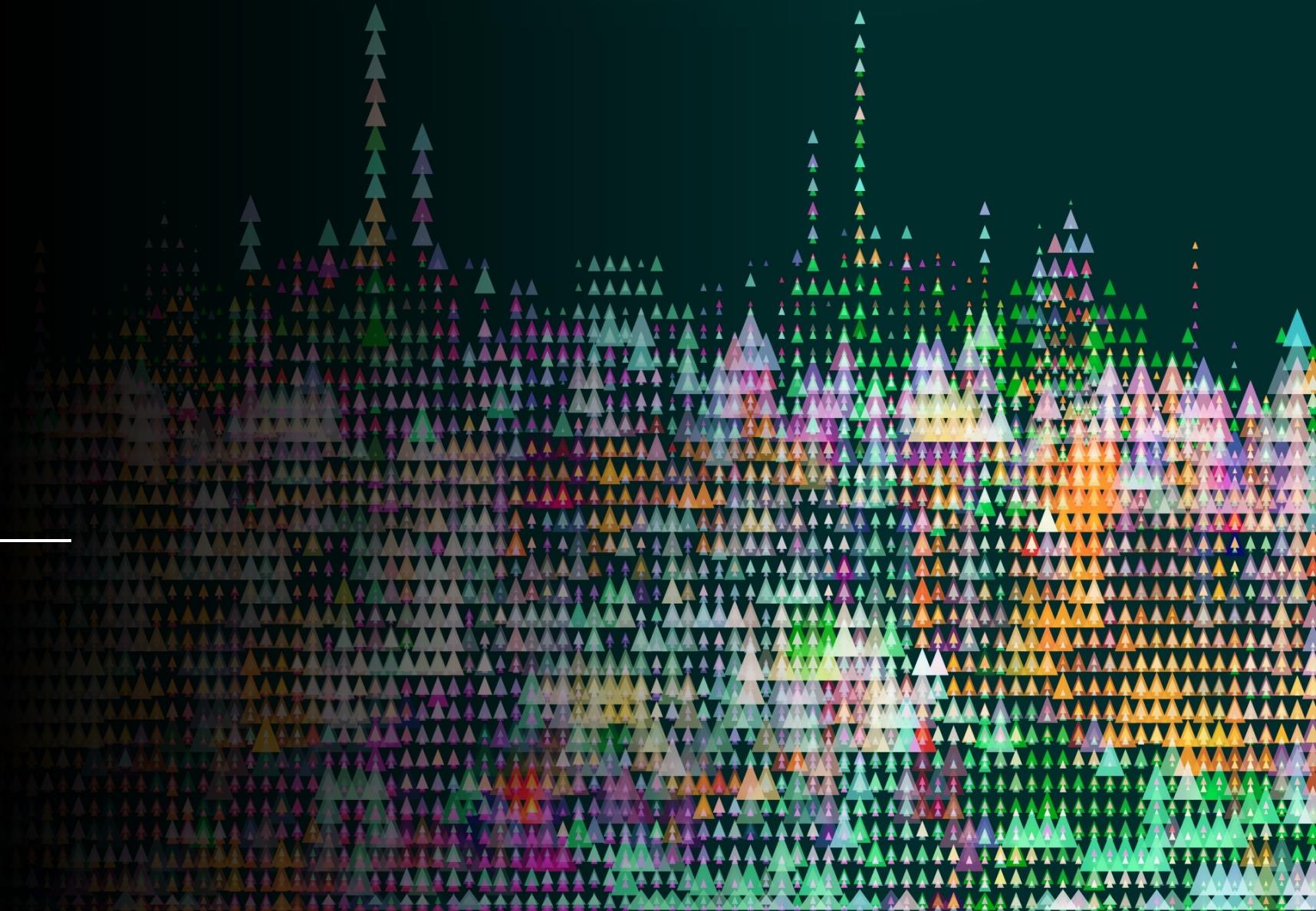




Descriptive Analytics through Data Storytelling

Pendampingan Data Storytelling

14/09/2023



Motivation



Bad graphs are everywhere,
including in places where you
would expect people to know
better



Proficiency in dashboard tools
is not enough, **your tools do
not know what is the story**
within your data



We learn a lot about language
and math, but we are not
naturally good in pairing them
(tell stories with numbers)



*We need an analyst or
communicator of the
information to bring the story
visually and contextually to life*

Reasons behind the need of visualization

- To evaluate what had just happened in the past
- To estimate current position, is it good? or compromising?
- To prepare the future better
- To monitor specific condition regularly
- To identify odd occurrences quickly
- *To communicate efficiently to the audience*



ITS
Institut
Teknologi
Sepuluh Nopember

ServiceDesk

HOME

CHECK TICKETS

KNOWLEDGE BASE

HOW TO CREATE TICKET

REGISTER AN ACCOUNT

❗ Pengumuman Penutupan Jam Layanan

Search our online knowledge base for answers ...

Need Support? Create a Ticket.

Ticket Subject

Your Email

Priority

 Low

Ticket type

Keluhan Permintaan/Pertanyaan

Unit of Destination

 Select ...

Chat with DPTSI ITS

Hi! How can we help you?

Member

zahra_

.....

Start chat

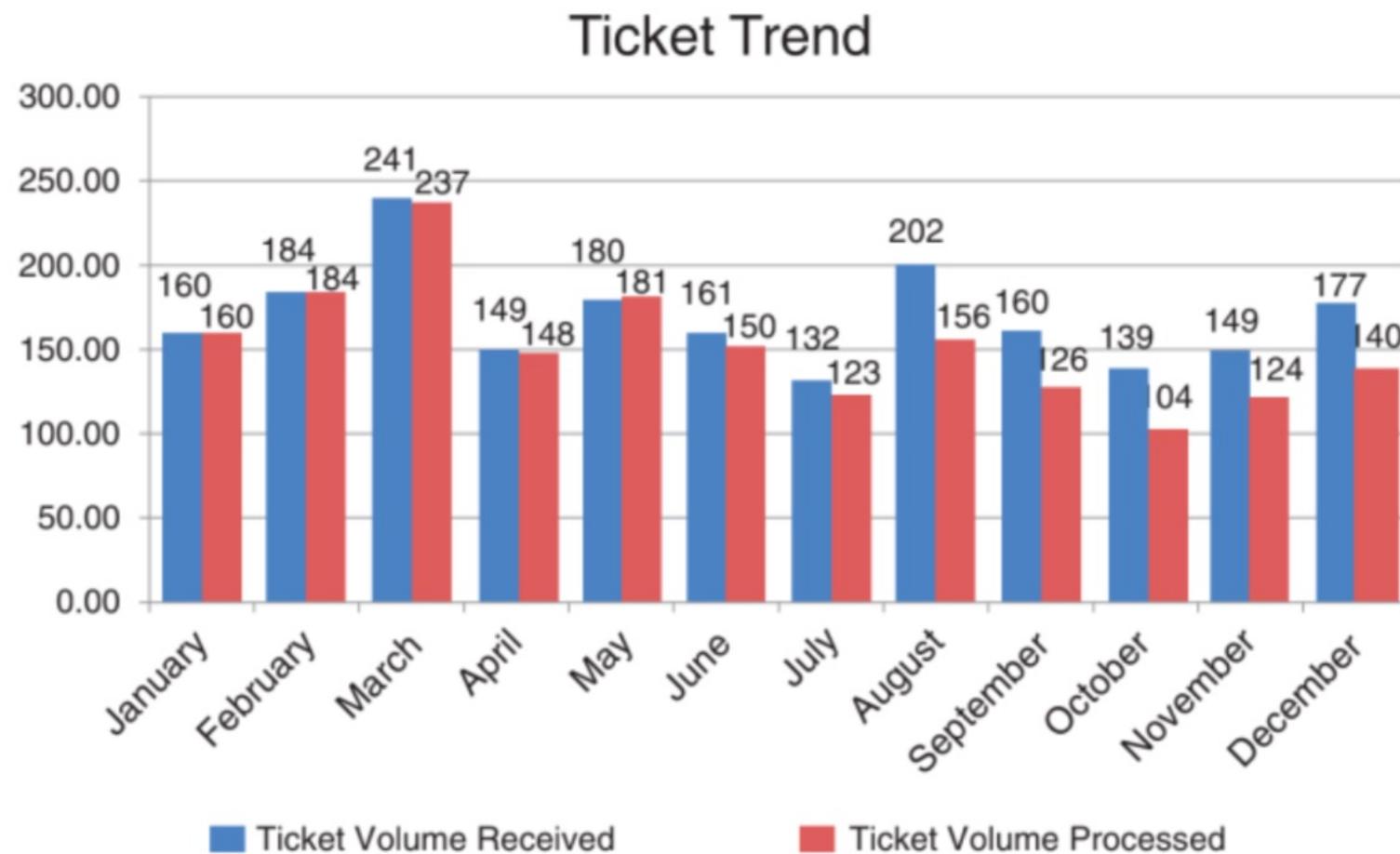
Powered by Messenger

Forgotten Pa...

Knowledge Base



Example 1



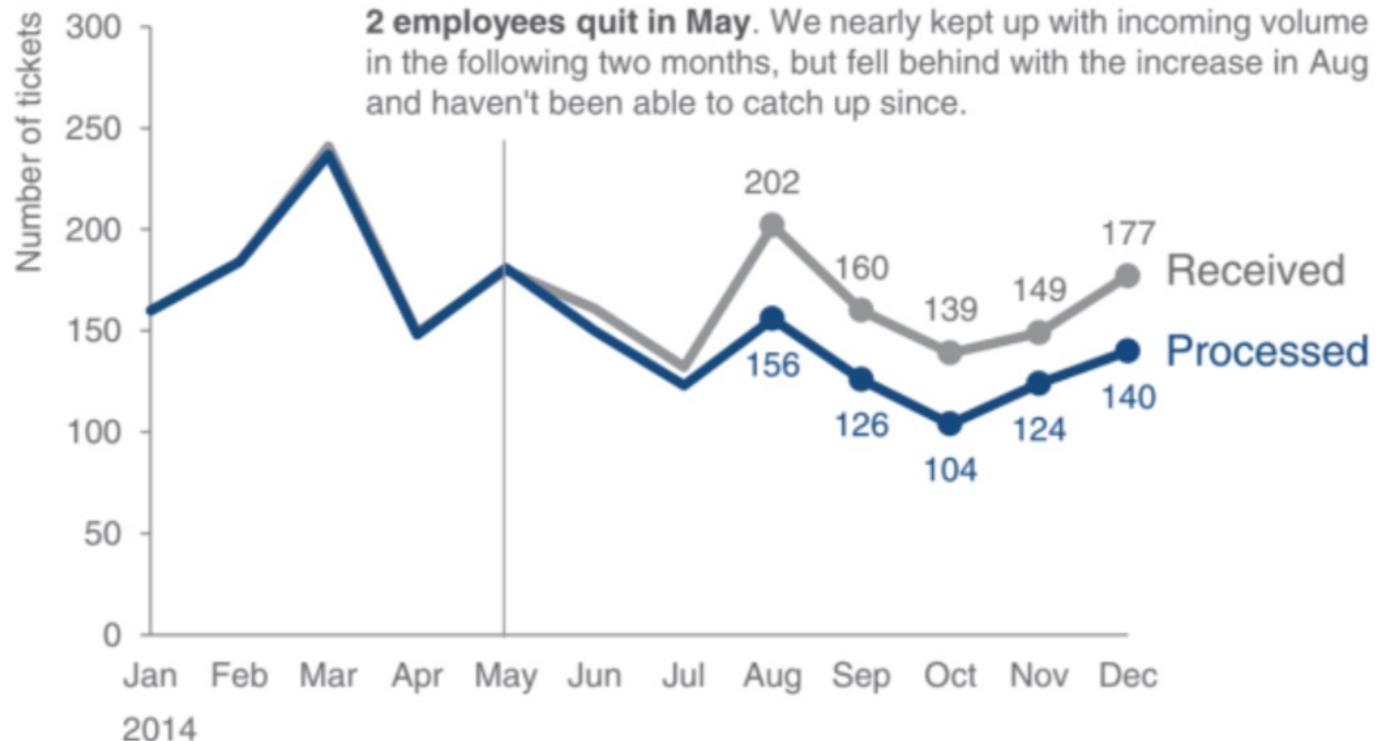
Example 1

*storytelling with
data instead of
showing data*

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

Highlight

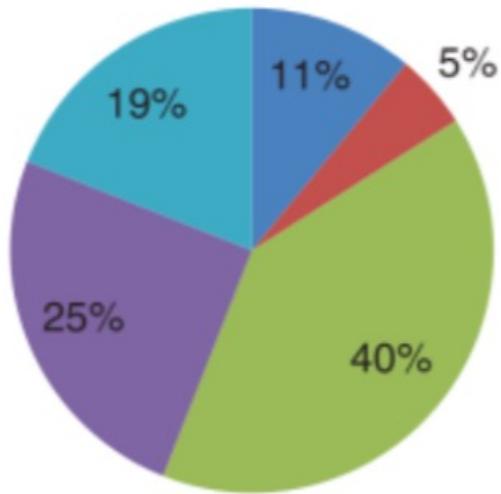
- Chart type
- Color: grey and blue
- Marker
- Showing differences without calculation
- Dividing graph into two section: before 2 FTEs leaving and after



Survey Results

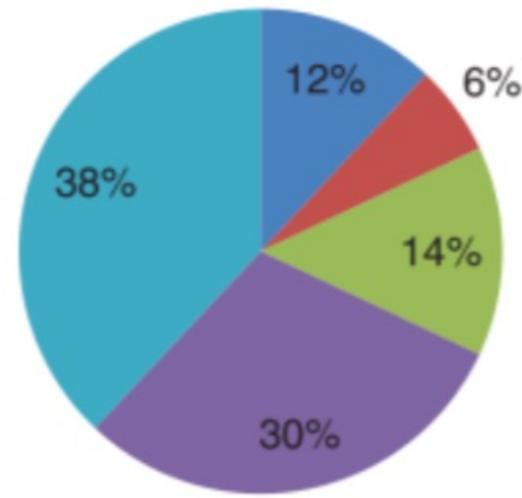
PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



Example 2

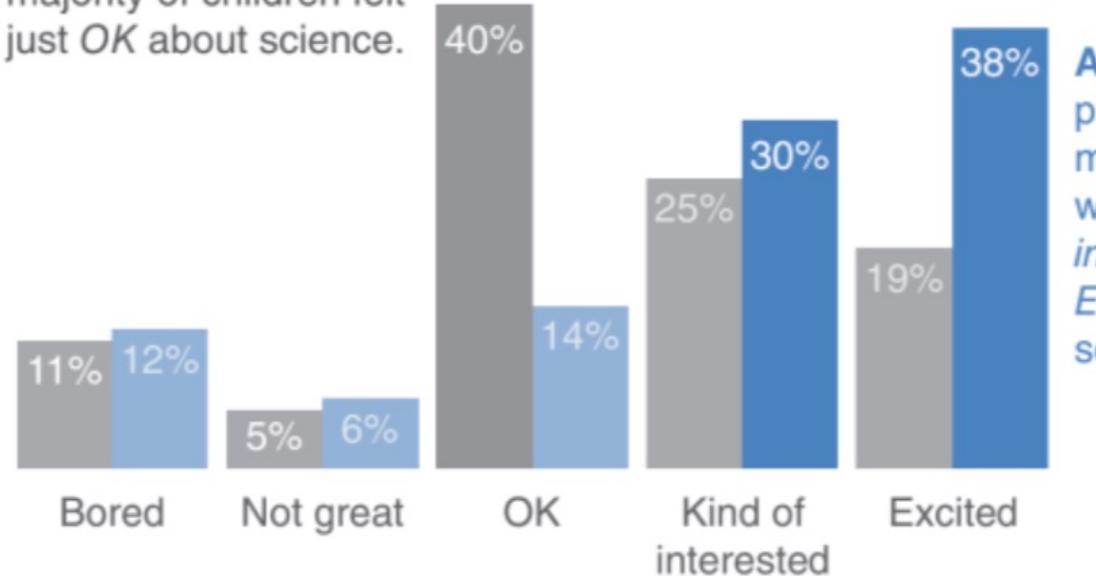
Example 2

*storytelling with
data instead of
showing data*

Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just *OK* about science.



AFTER program,
more children
were *Kind of
interested &
Excited* about
science.

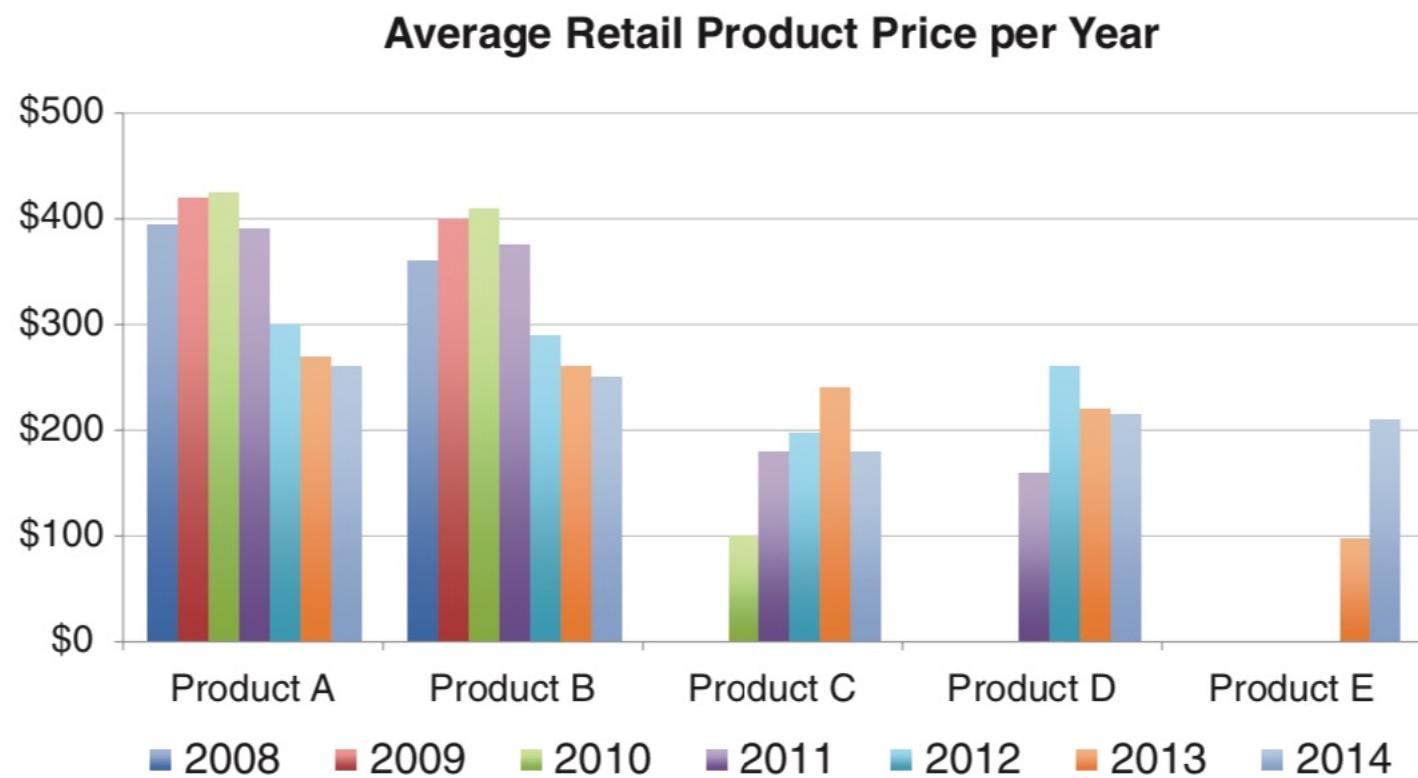
Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

Highlight

- Showing preferences from left to right, consistent with the scale
- Darker color showing good outcomes
- Color depends on the audiences too



Example 3

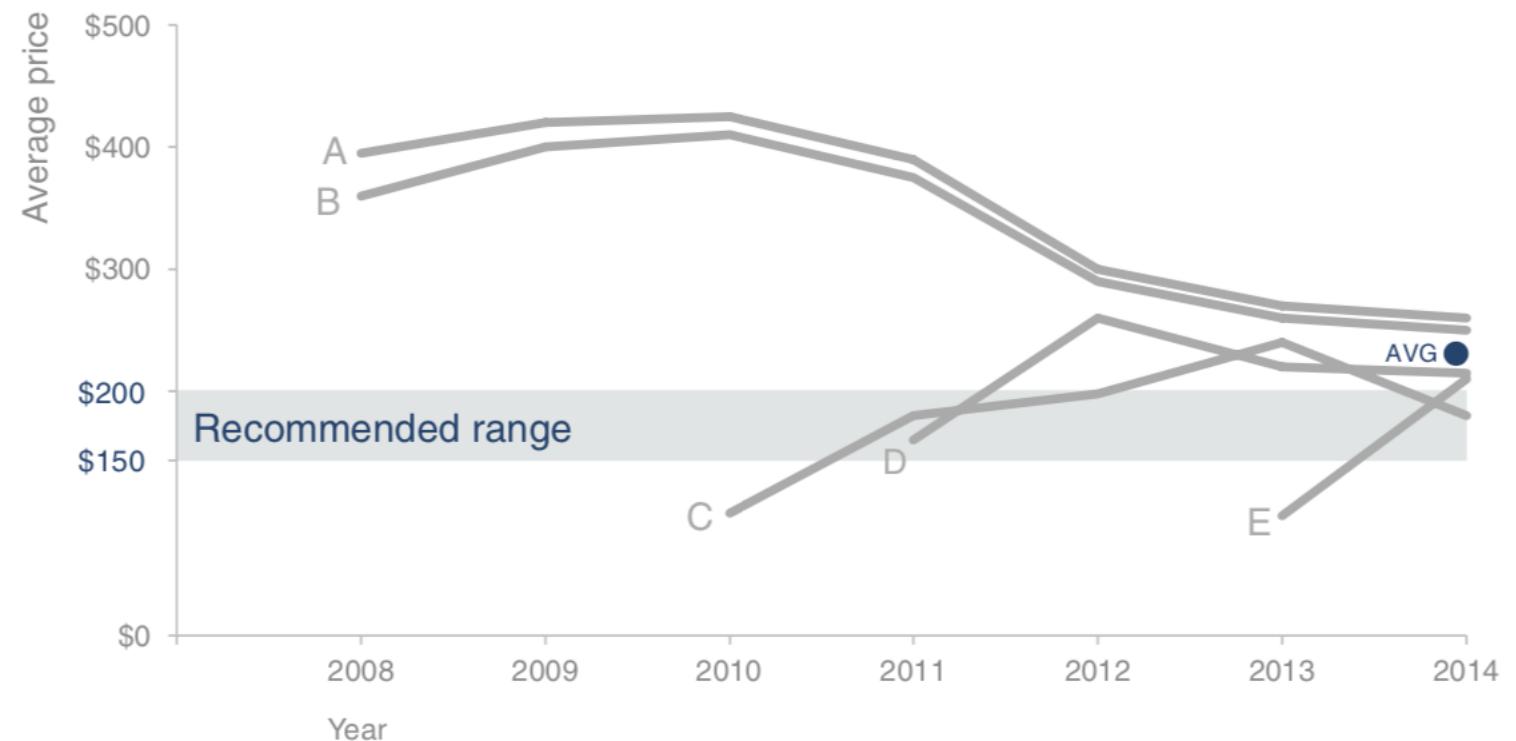


Example 3

*storytelling with
data instead of
showing data*

To be competitive, we recommend introducing our product *below* the \$223 average price point in the **\$150–\$200 range**

Retail price over time by product



Highlight

- Different starting point: not apple-to-apple for each product
- Helping future decision making on price of new product
- Put average value

How to tell stories with data: 6 lessons

1

Understand the context

2

Choose an appropriate visual display

3

Eliminate clutter

4

Focus attention where you want it

5

Think like a designer

6

Tell a story

Choose an effective visual

Simple text

Tables Heatmap

Points (Scatterplot)

Lines

Slopegraph

Bars Vertical & Horizontal bar chart Stacked bar chart Waterfall chart

Area

Tips: Specific graph types and elements you should avoid

Simple Text

The fact that you have some numbers does not mean that you need a graph!

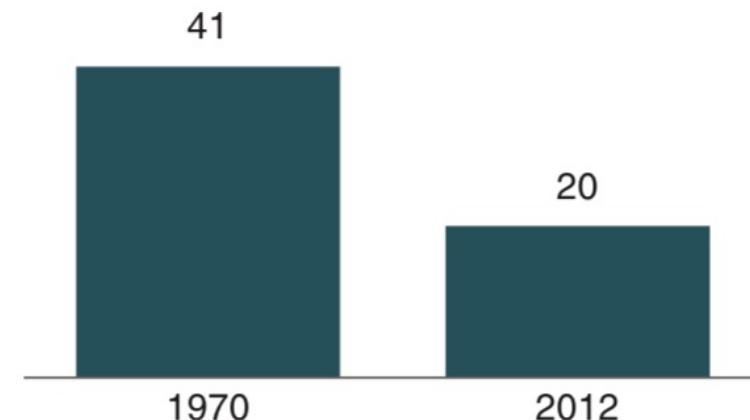
When you have a number or two that you want to communicate, think about solely using the number—making it as prominent as possible—and a few supporting words to clearly make your point.

20%

of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970

Children with a "Traditional" Stay-at-Home Mother

% of children with a married stay-at-home mother with a working husband



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

Simple Text

- Bapak Prof. Gading Ekapuja memperoleh peringkat top 1% dari Scopus publication.

Tables

The data should be what stands out, not the borders.

Using a table in a live presentation is rarely a good idea ,
consider including the full table in the appendix and a link or reference.

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Heatmap

A heatmap is a way to visualize data in tabular format, where in place of (or in addition to) the numbers, you leverage colored cells that convey the relative magnitude of the numbers.

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW-HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

the lowest number (11%) and highest number (58%)

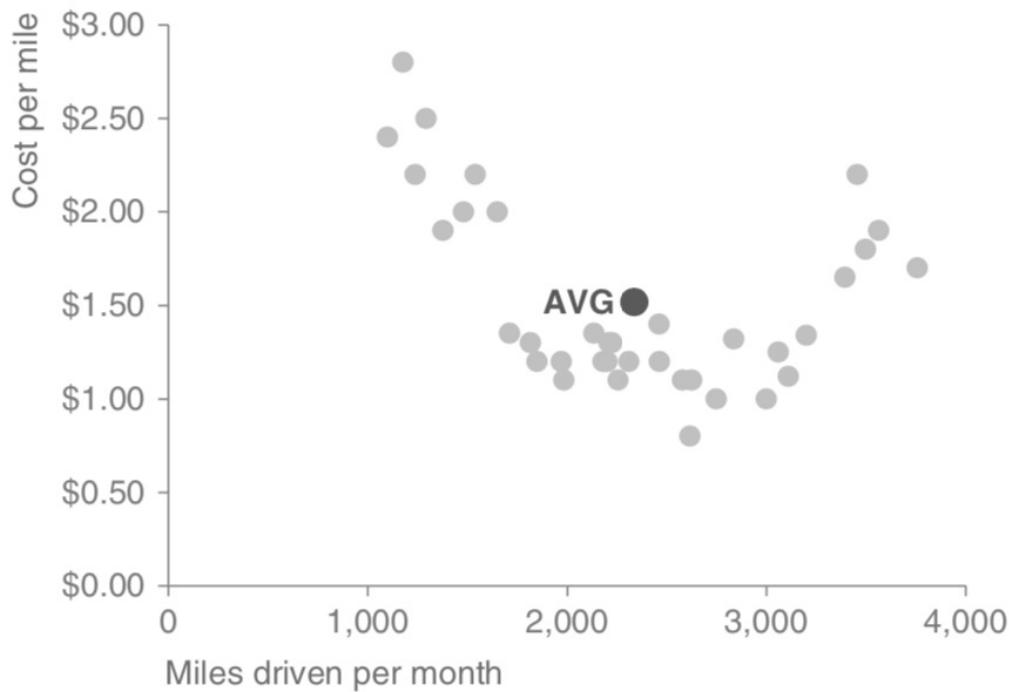
Heatmap

- Tingkat keketatan prodi (audience: ortu camaba)
 - Dark untuk prodi yang paling ketat
 - Light untuk prodi yang paling longgar
- Rata-rata lama studi per fakultas (audience: Bidang Akademik)
 - Dark untuk fakultas yang on time sesuai jenjang studi
 - Light untuk fakultas yang melebihi standar

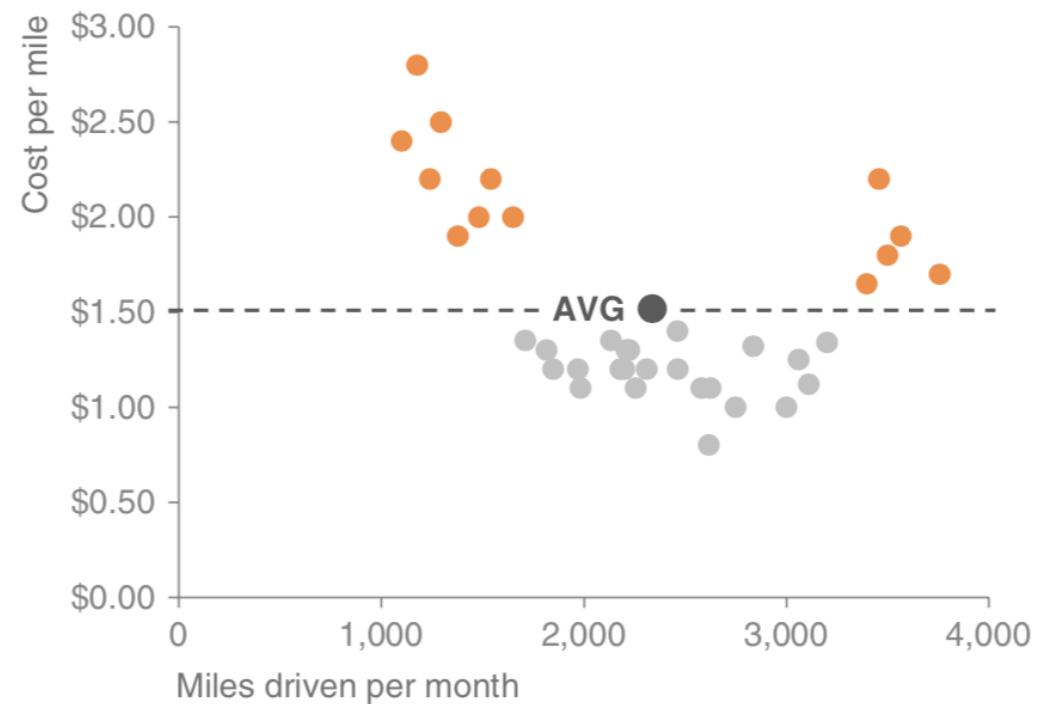
Scatterplot

“chart” is the broader category, with “graphs” being one of the subtypes (other chart types include maps and diagrams)

Cost per mile by miles driven

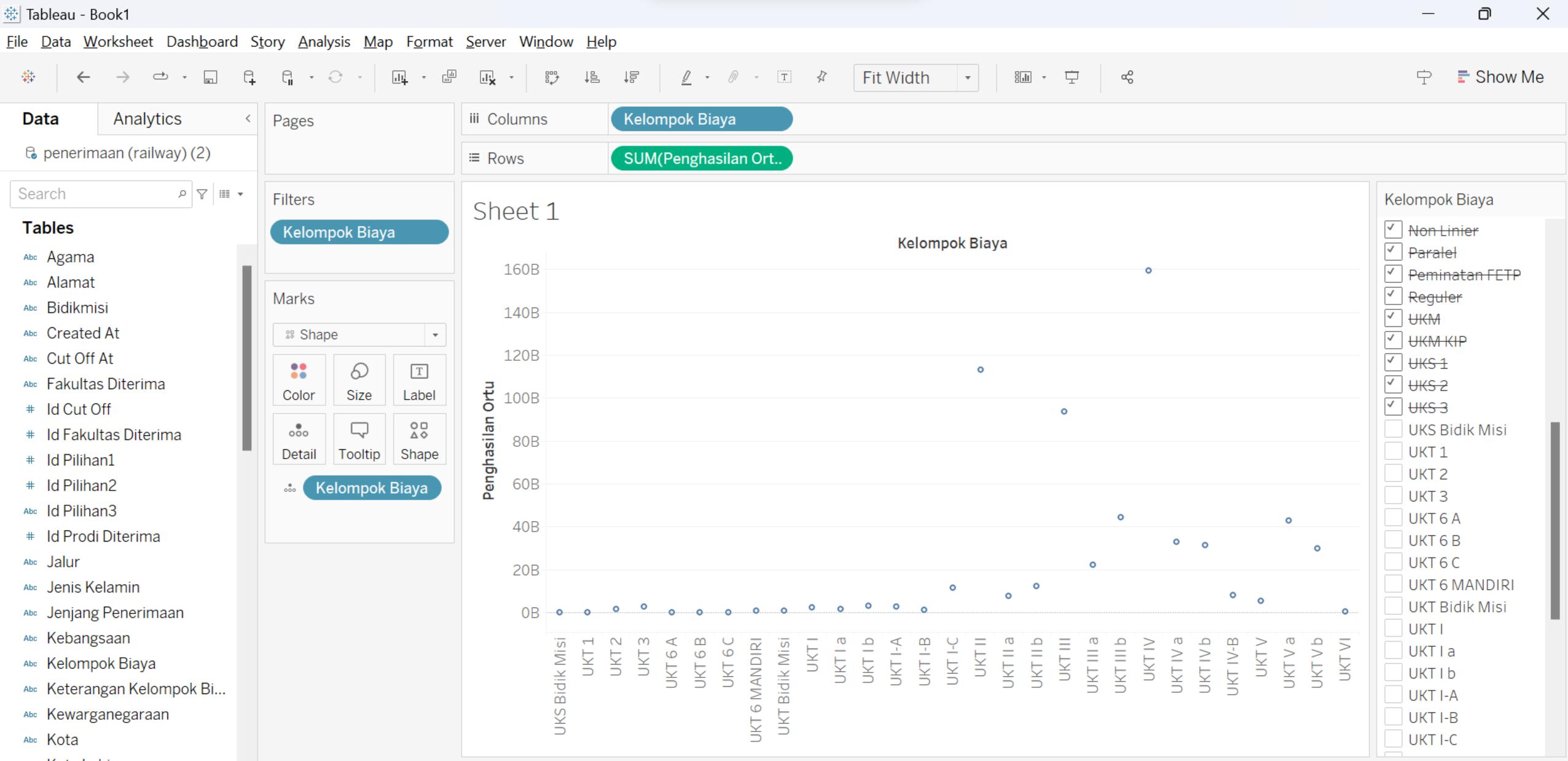


Cost per mile by miles driven



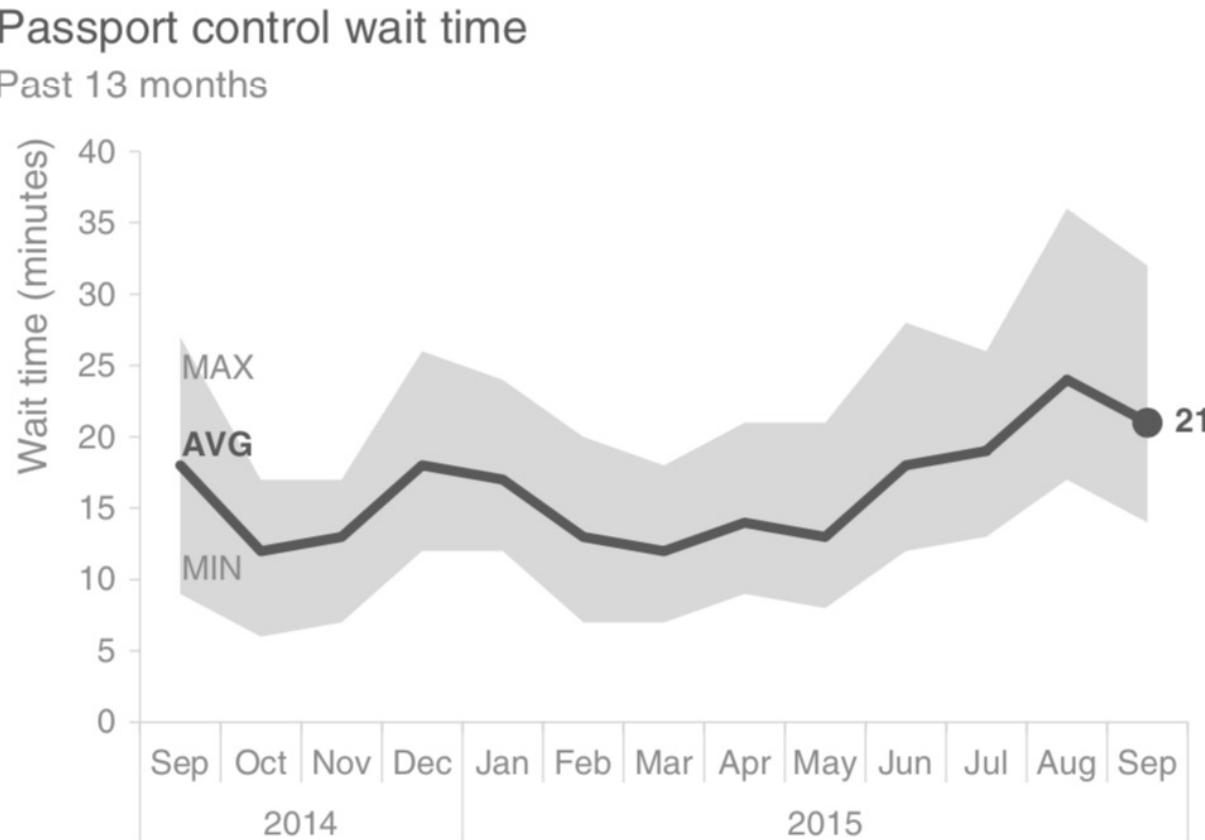
Scatterplot

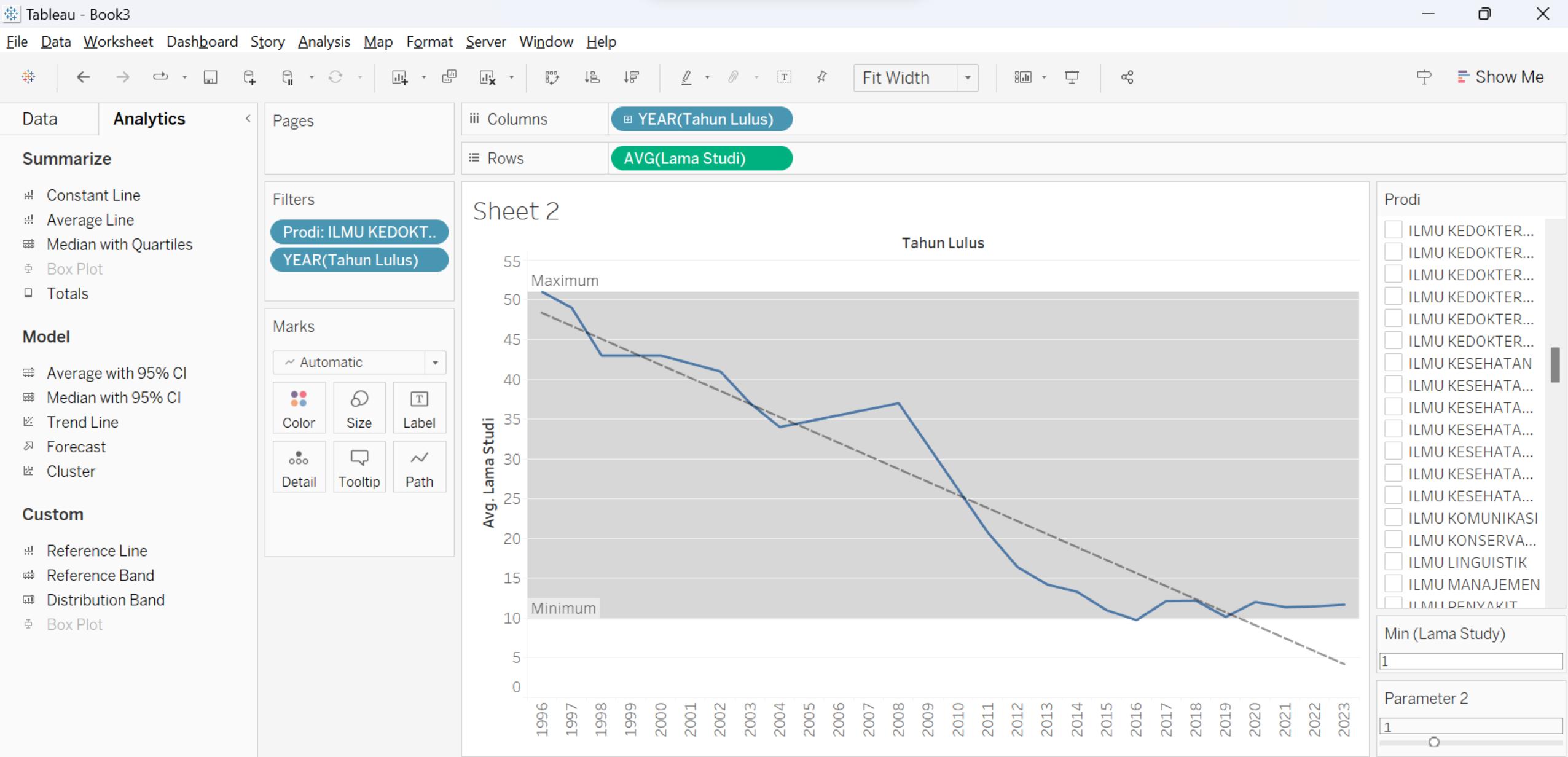
- Bidang Keuangan ingin menganalisa apakah terdapat anomaly pada UKT. Contoh anomaly yang dimaksud adalah
 - Orang tua dengan penghasilan tinggi mendapatkan UKT yang rendah
 - Orang tua dengan penghasilan rendah mendapatkan UKT yang tinggi
- Asumsi
 - UKT I paling rendah
 - UKT V paling tinggi



Lines

The line graph can show a single series of data, two series of data, or multiple series. Note that when you're graphing time on the horizontal x-axis of a line graph, the data plotted must be in consistent intervals (daily, weekly, monthly, yearly).





Data Source

Sheet 1

Sheet 2



23 marks 1 row by 28 columns SUM of AVG(Lama Studi): 586.32

14:34

16/09/2023 3

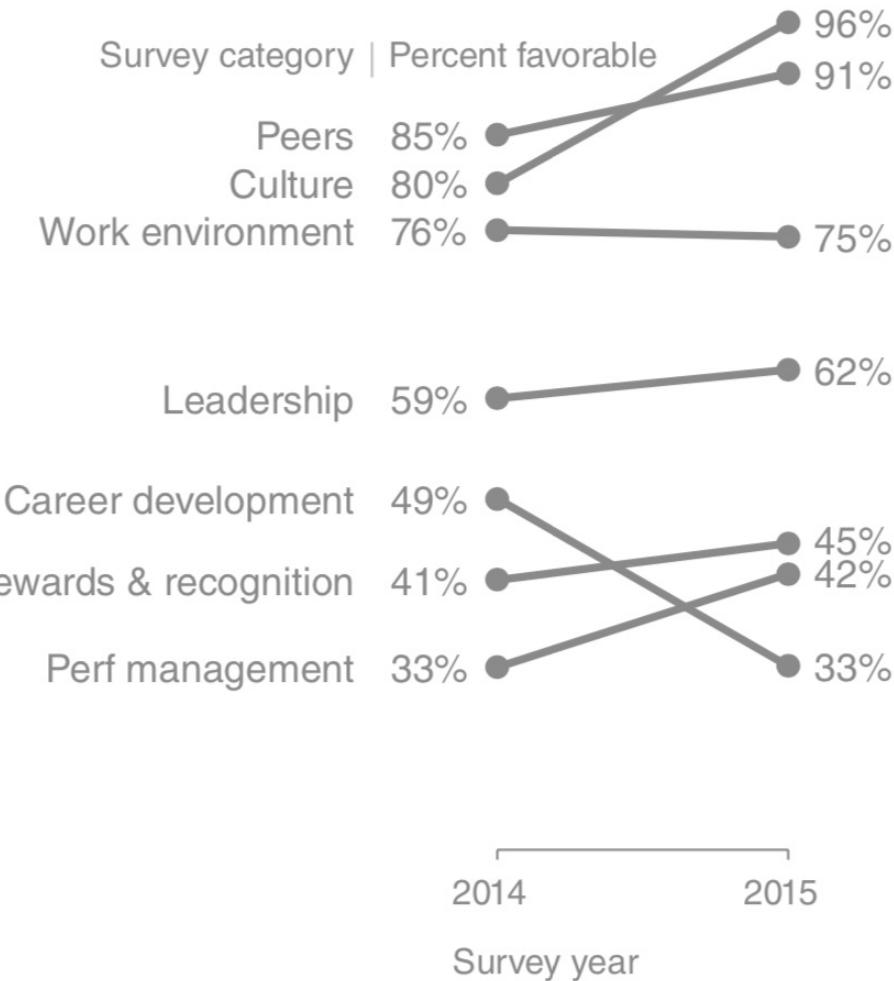
31°C
Cerah

Search

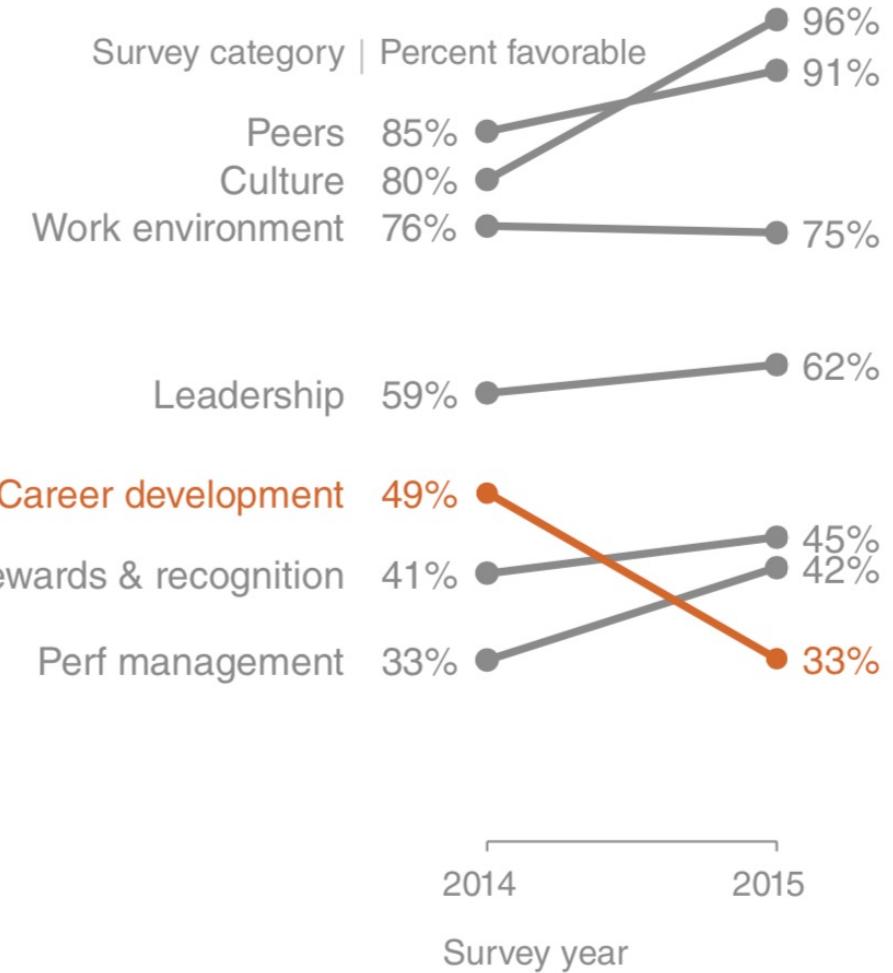


Slopegraph

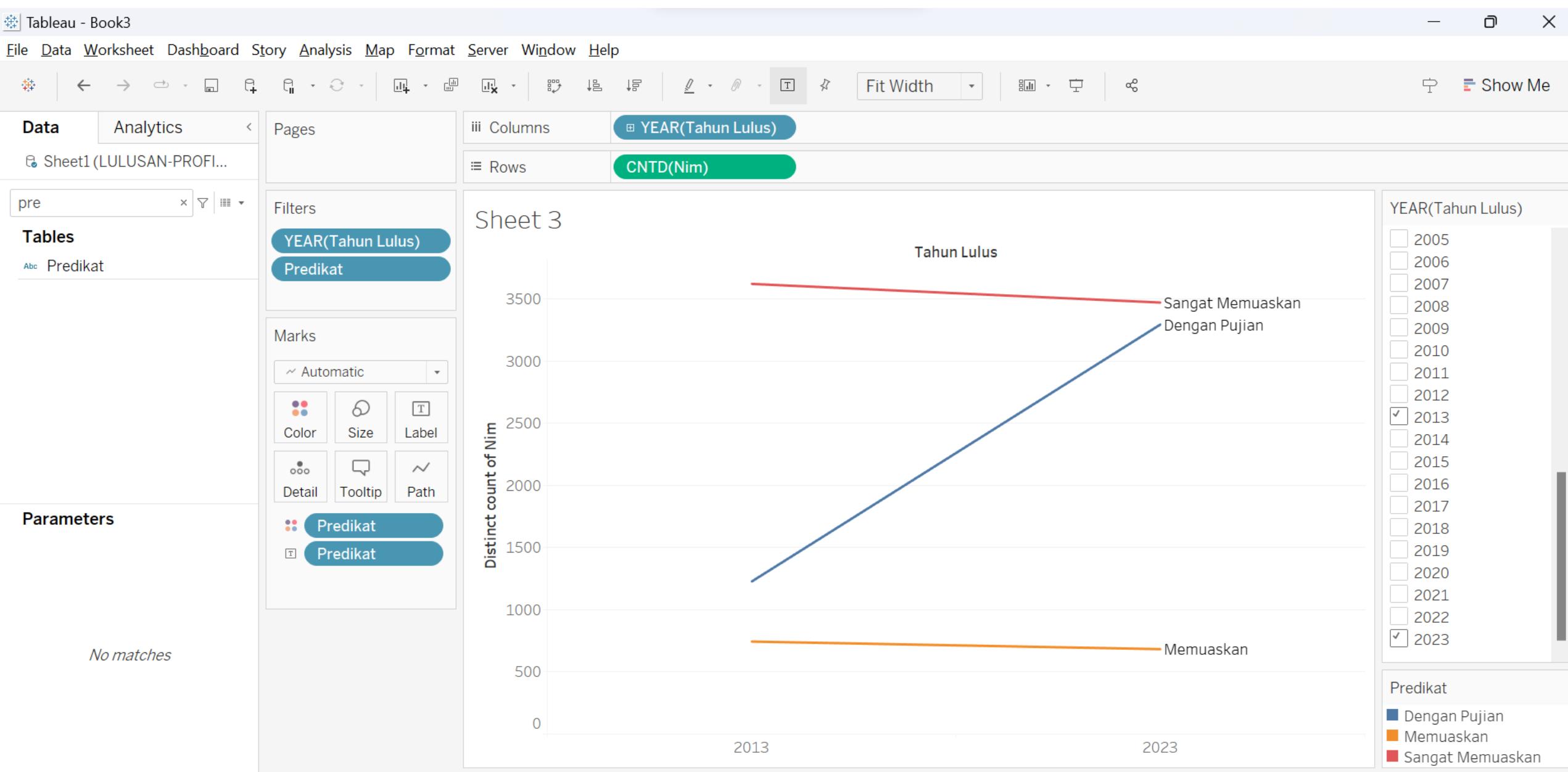
Employee feedback over time



Employee feedback over time



- Bidang Akademik ingin mengetahui apakah anggapan “inflasi” IPK itu memang benar terjadi pada 10 tahun terakhir?



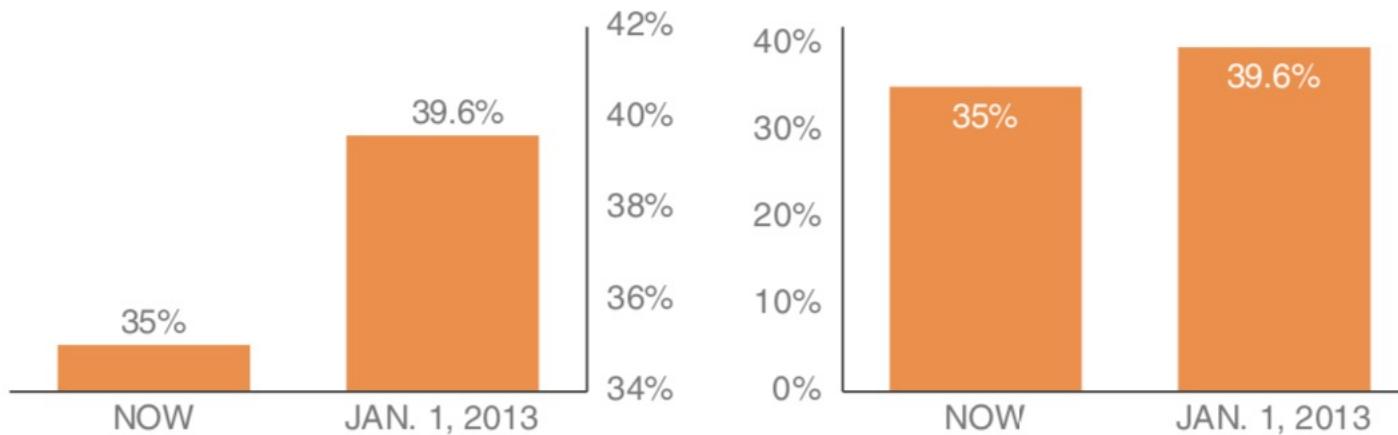
Bars

Sometimes bar charts are avoided *because they are common*. This is a mistake. Rather, bar charts should be leveraged, as this means less of a learning curve for your audience.

Bar charts must have a zero baseline

Non-zero baseline: as originally graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

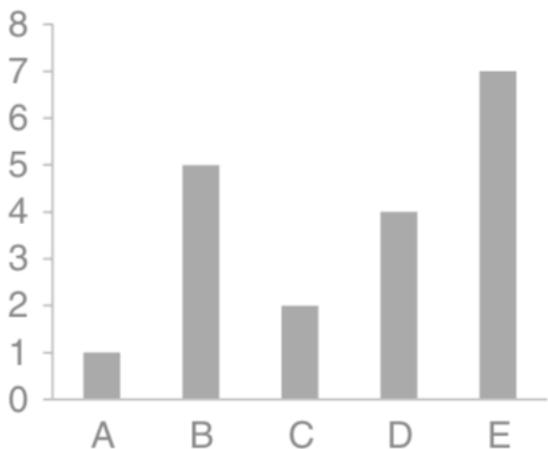


the visual increase is 460% (the heights of the bars are $35 - 34 = 1$ and $39.6 - 34 = 5.6$, so $(5.6 - 1)/1 = 460\%$)

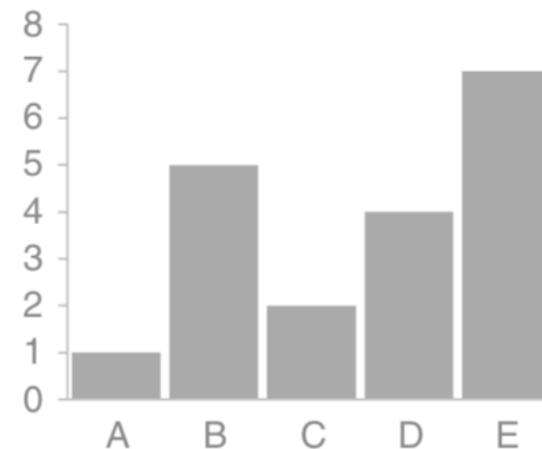
an actual visual increase is 13% from $((39.6 - 35) / 35)$

Width of the bar

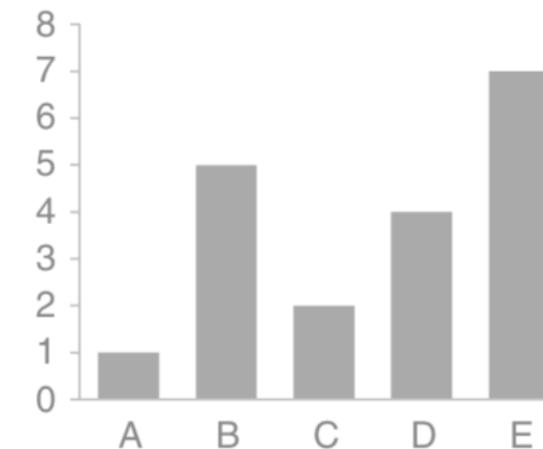
Too thin



Too thick



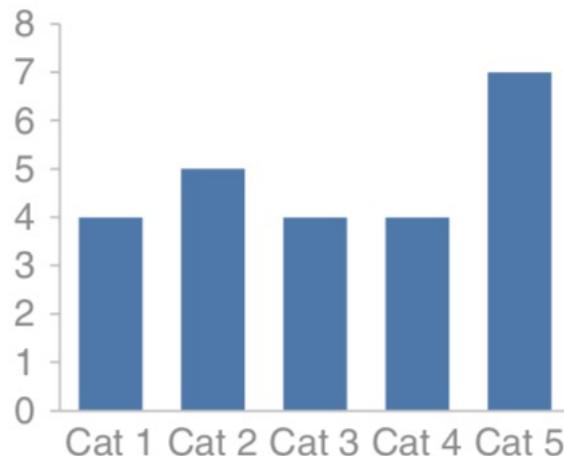
Just right



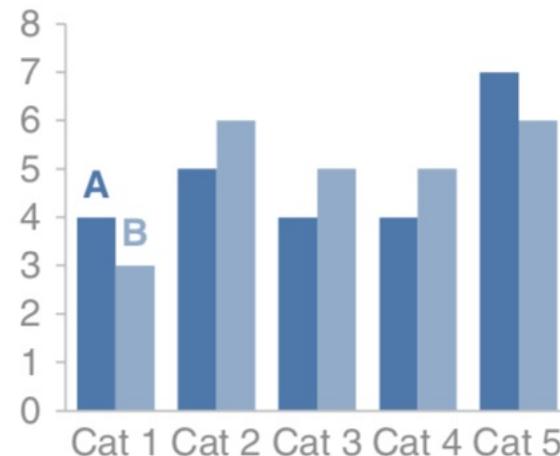
Vertical bar chart

Note that as you add more series of data, it becomes more difficult to focus on one at a time and pull out insight, so use multiple series bar charts with caution.

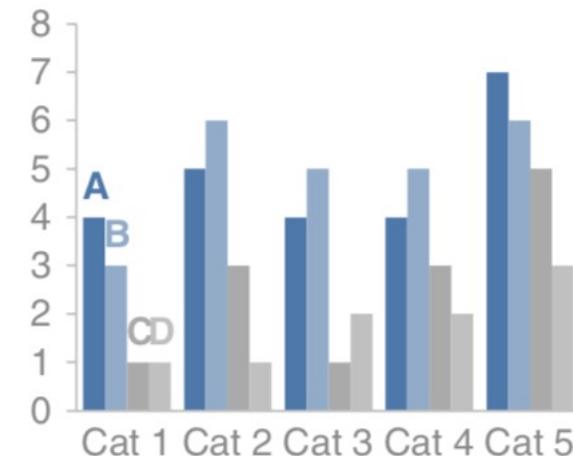
Single series



Two series



Multiple series



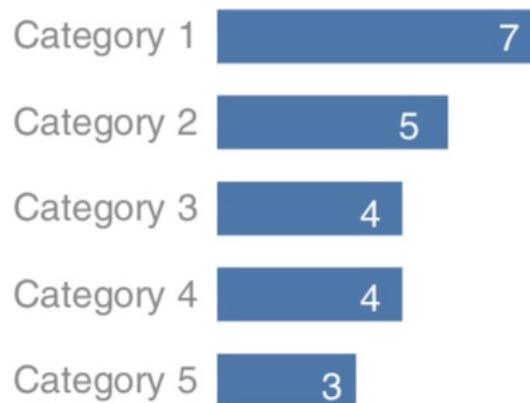
Horizontal bar chart

If we had to pick a single go-to graph for categorical data, it would be the horizontal bar chart.

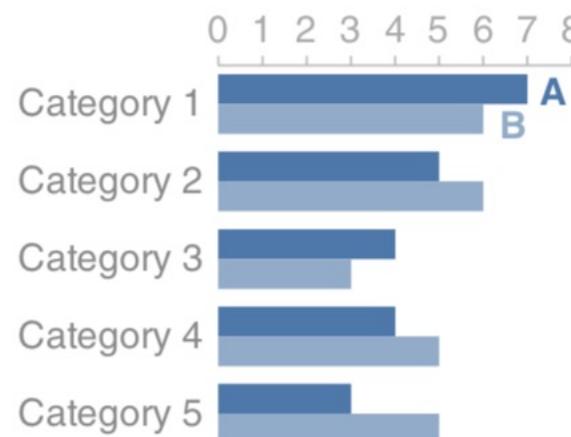
The horizontal bar chart is especially useful if your **category names are long**,

The structure of the horizontal bar chart is such that our eyes hit the category names before the actual data. This means by the time we get to the data, we already know what it represents.

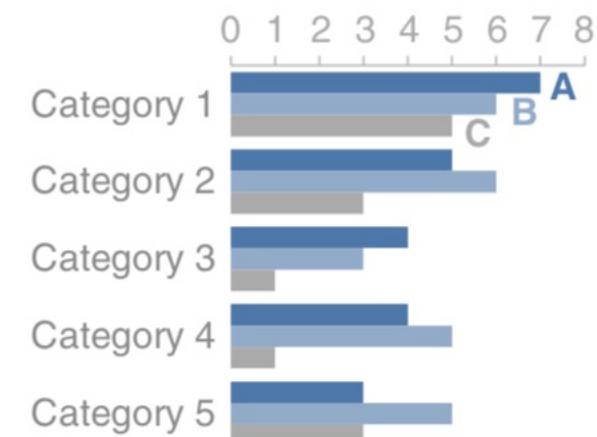
Single series



Two series

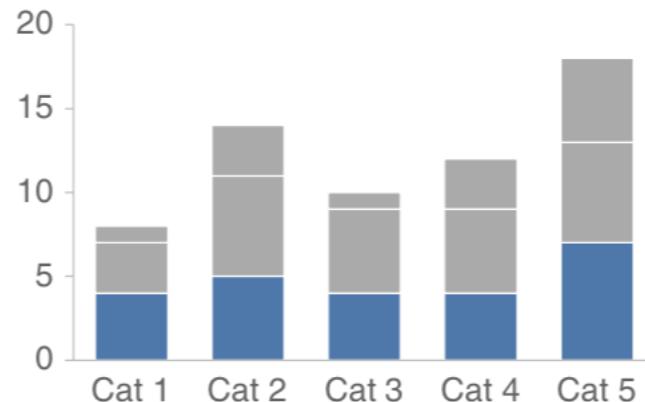


Multiple series

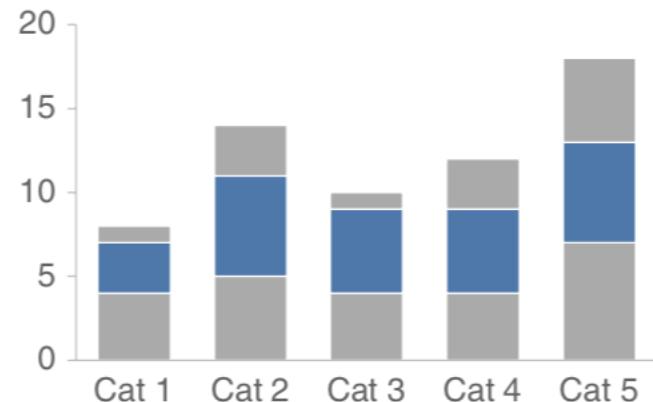


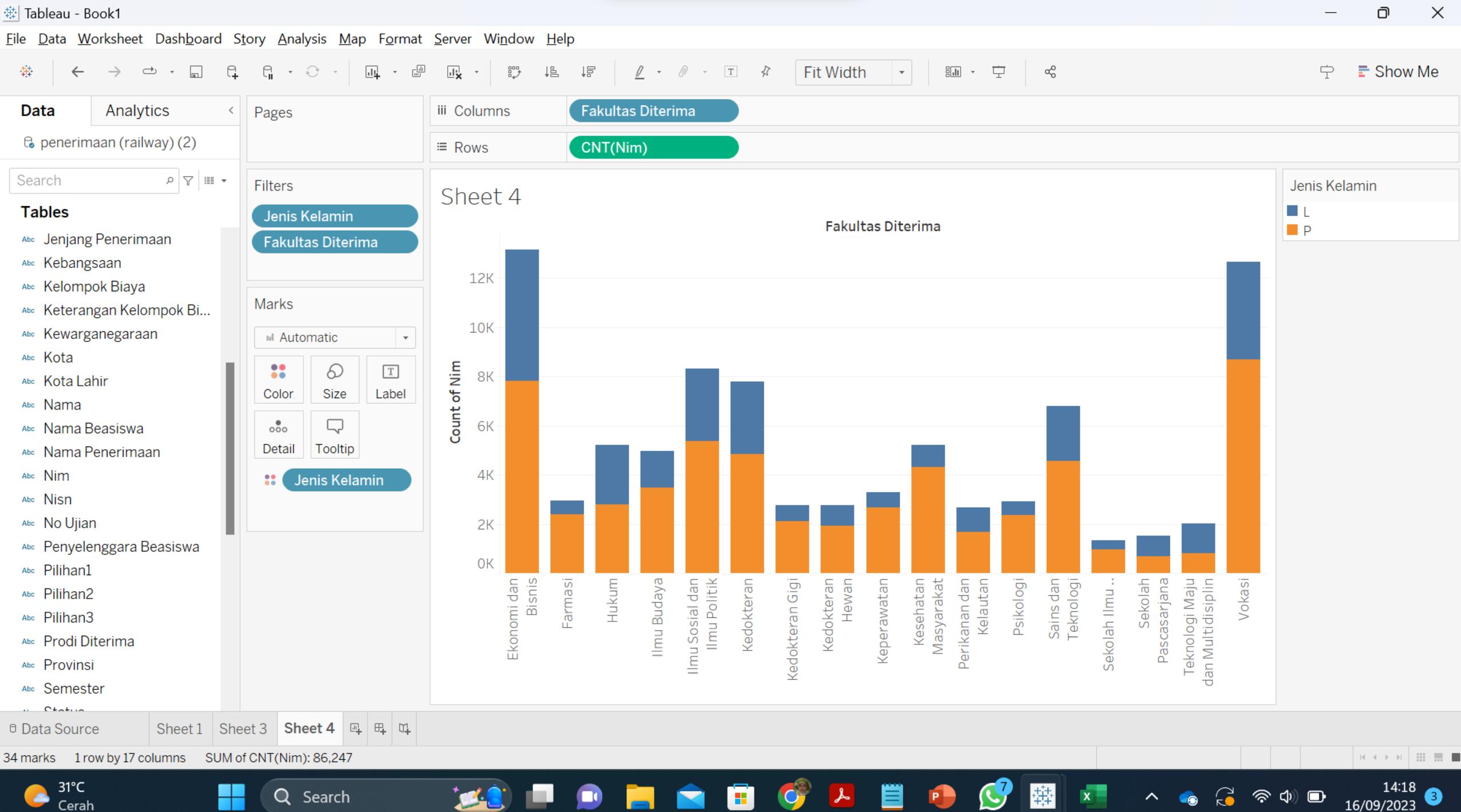
Stacked vertical bar chart

Comparing **these** is easy



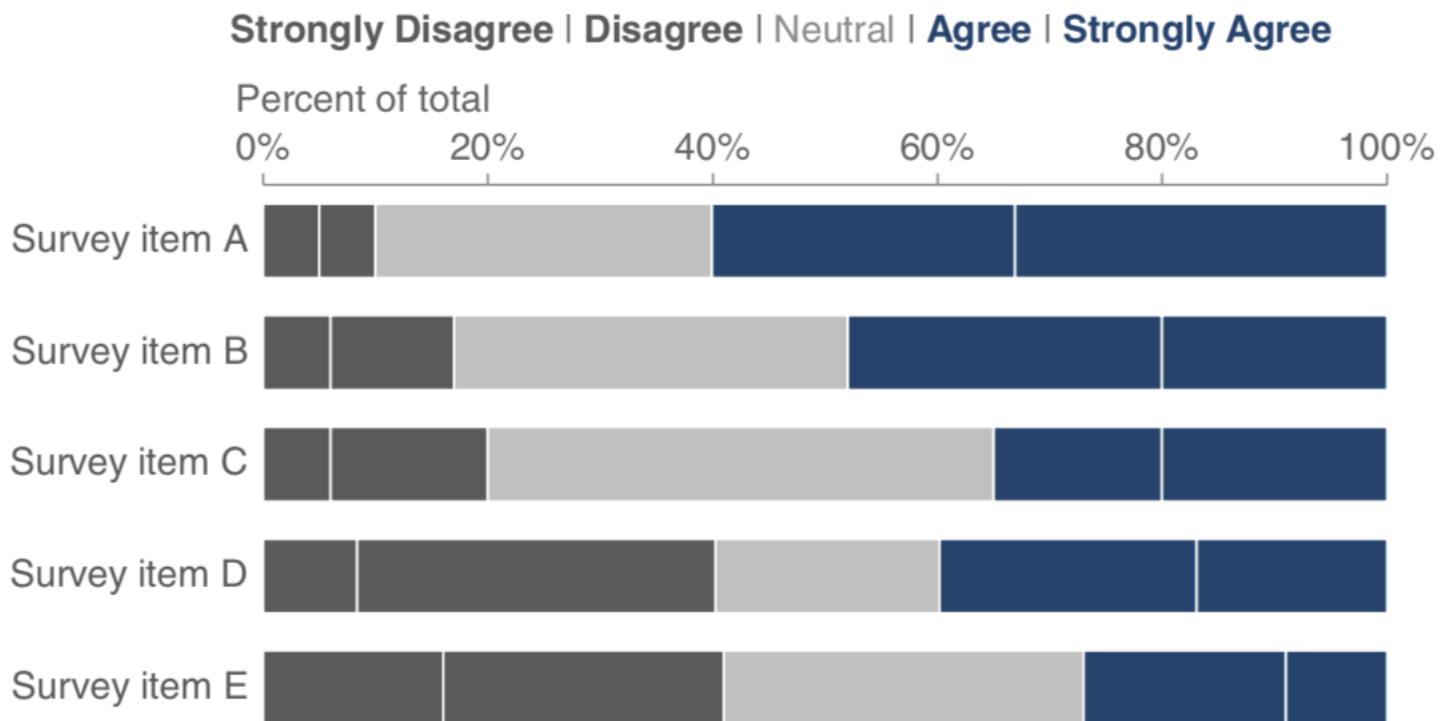
Comparing **these** is hard





Stacked horizontal bar chart

Survey results

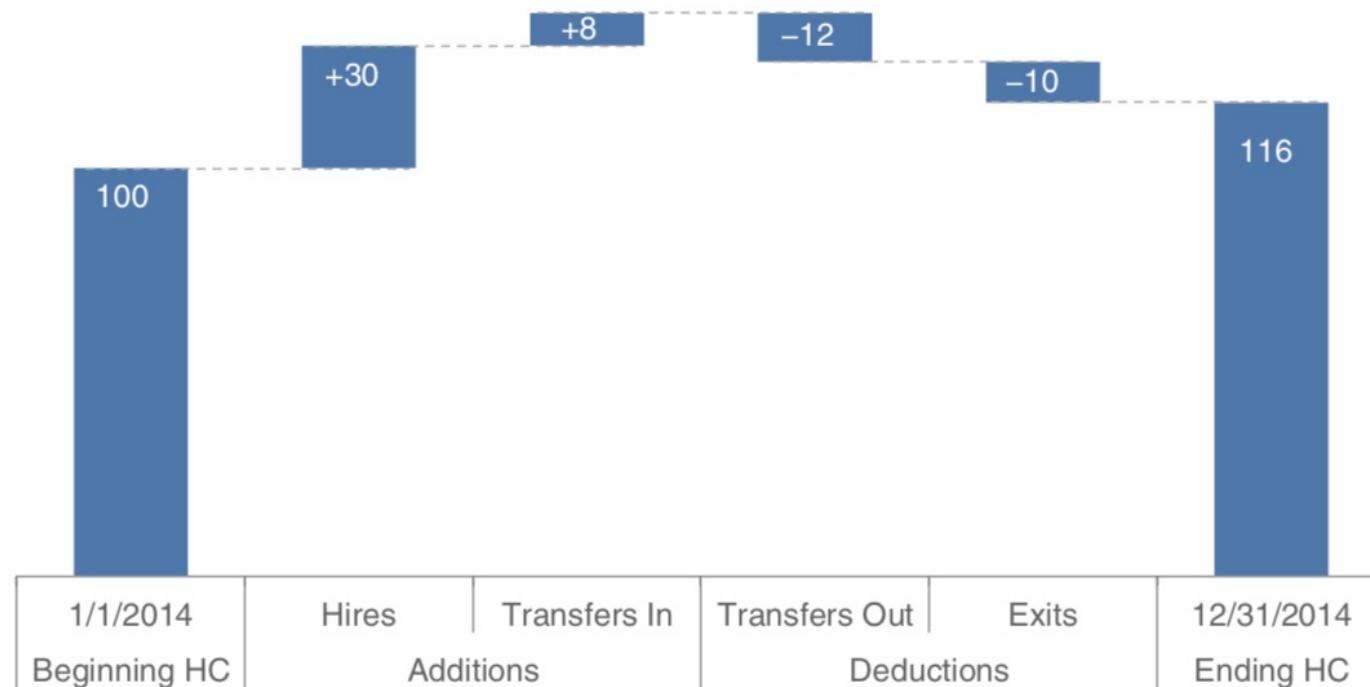


Waterfall chart

The waterfall chart can be used to pull apart the pieces of a stacked bar chart to focus on one at a time, or to show a starting point, increases and decreases, and the resulting ending point.

2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.

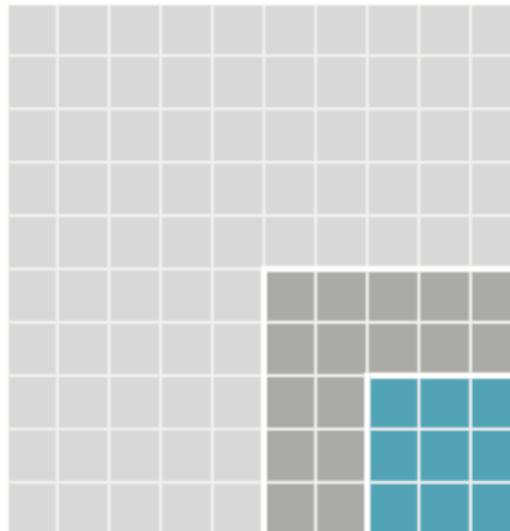


Area

Use it when you need to visualize numbers of vastly different magnitudes.

Typically avoid. Humans' eyes don't do a great job of attributing quantitative value to two-dimensional space, which can render area graphs harder to read than some of the other types of visual displays.

Interview breakdown



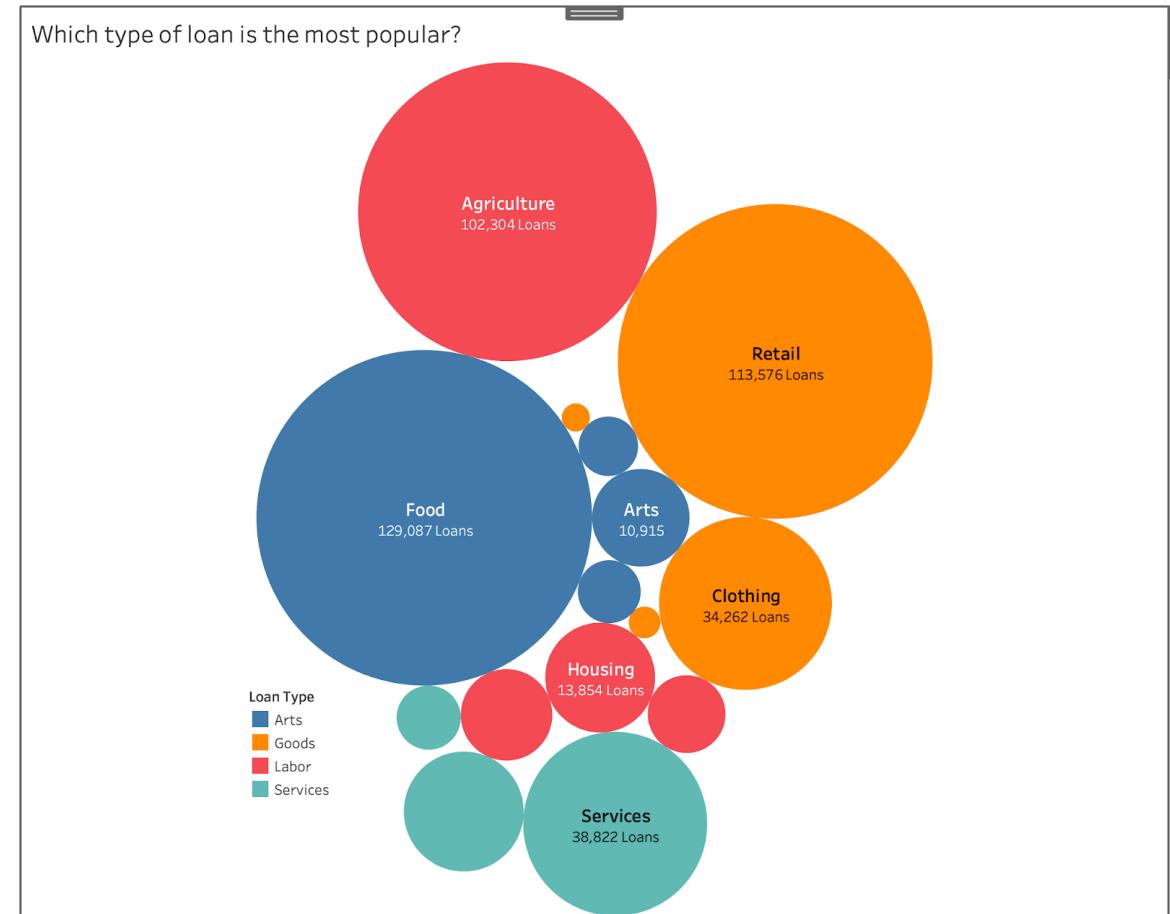
Out of every 100
phone screens...

we bring **25**
candidates onsite
for interviews...

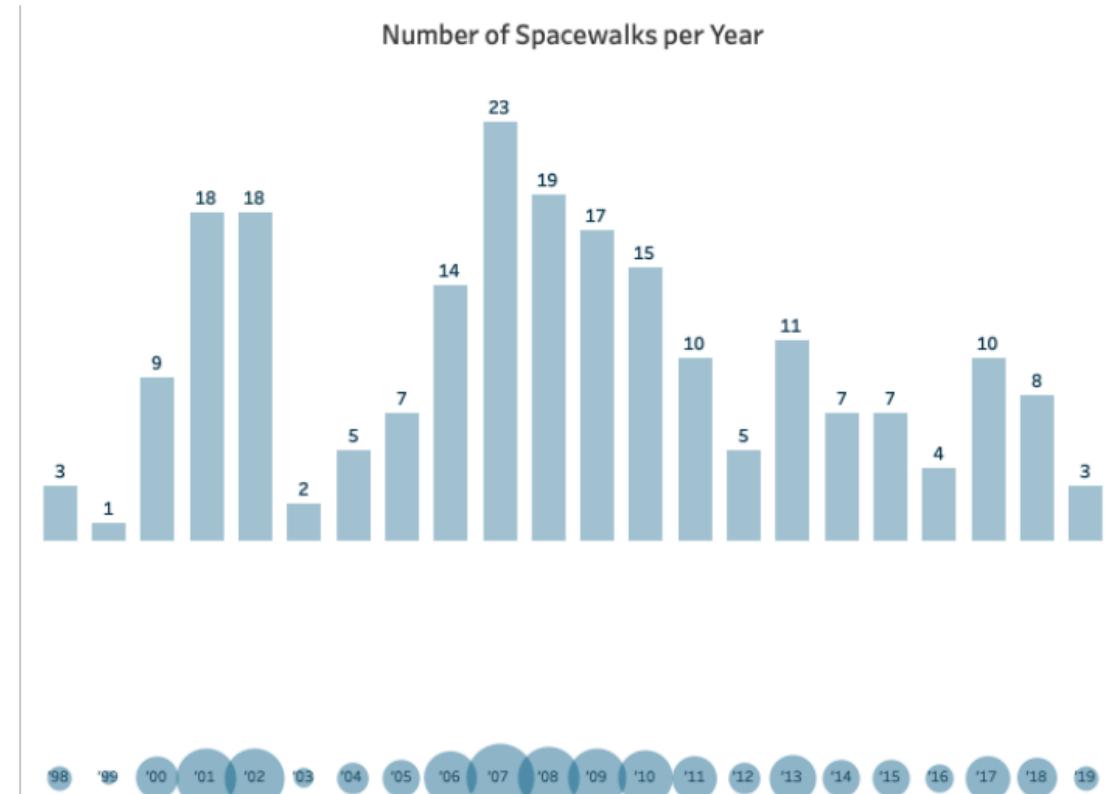
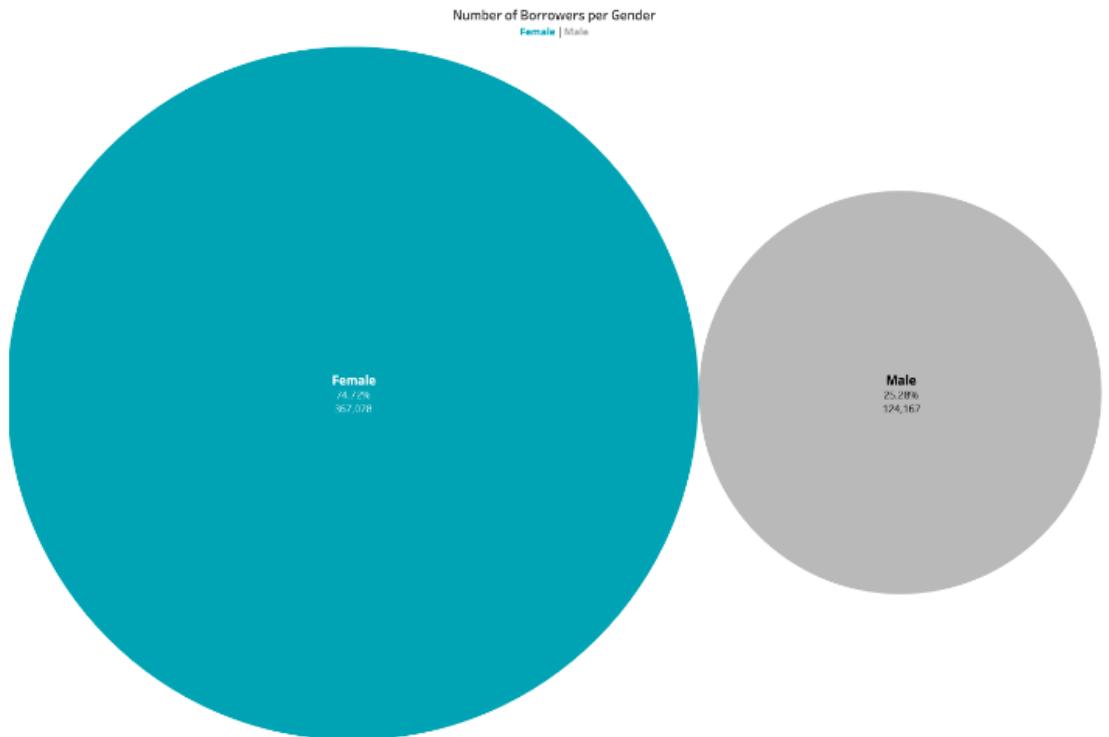
and
extend 9 offers.

Bubble Chart

- Bubble charts can help to simplify a complex dataset into a visual that makes it easy for viewers to **compare members** within a given field.
- Color then categorizes the bubbles by the type of spend they represent.



Great Examples of Bubble Charts



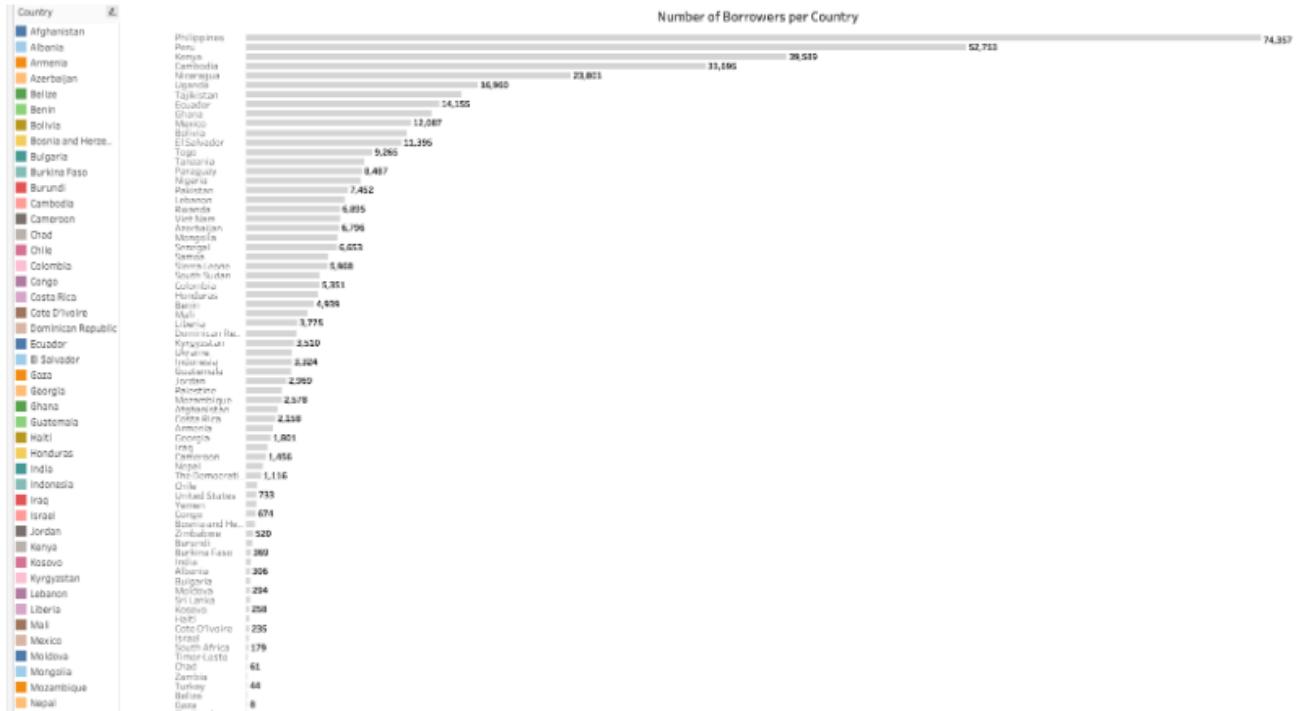
Bad Examples of Bubble Charts and Alternatives

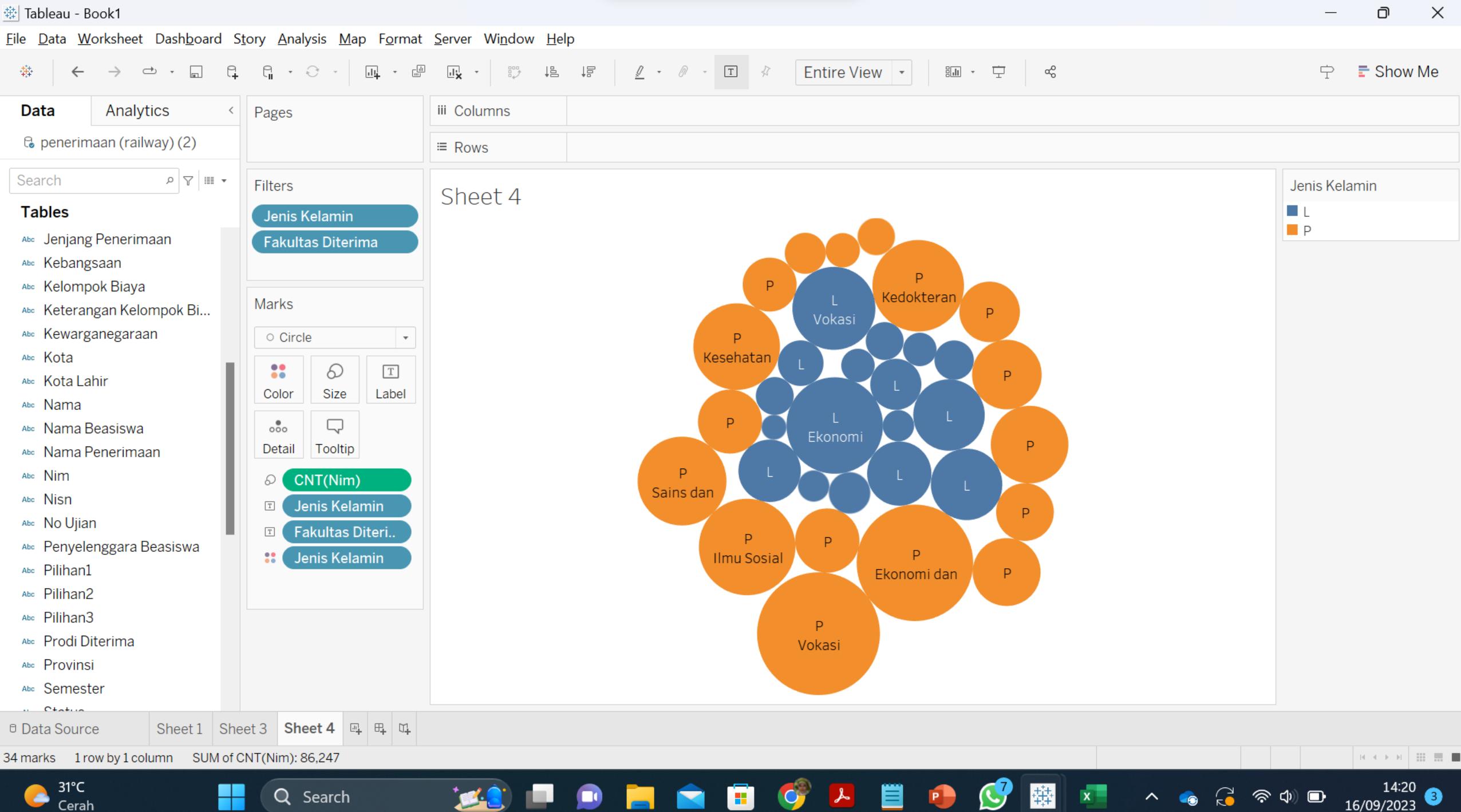


POOR EXAMPLE

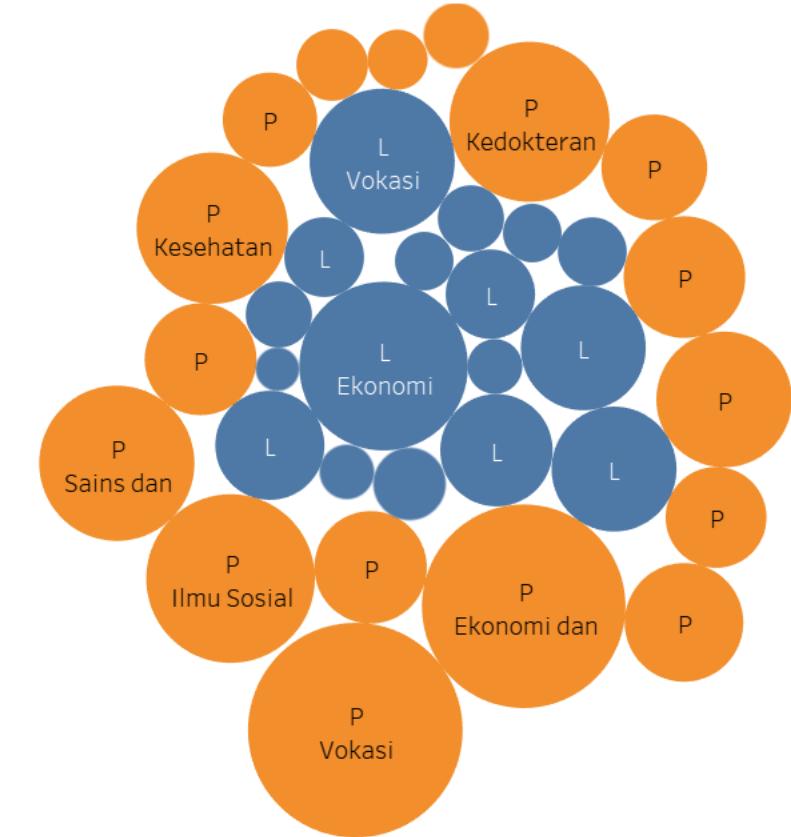
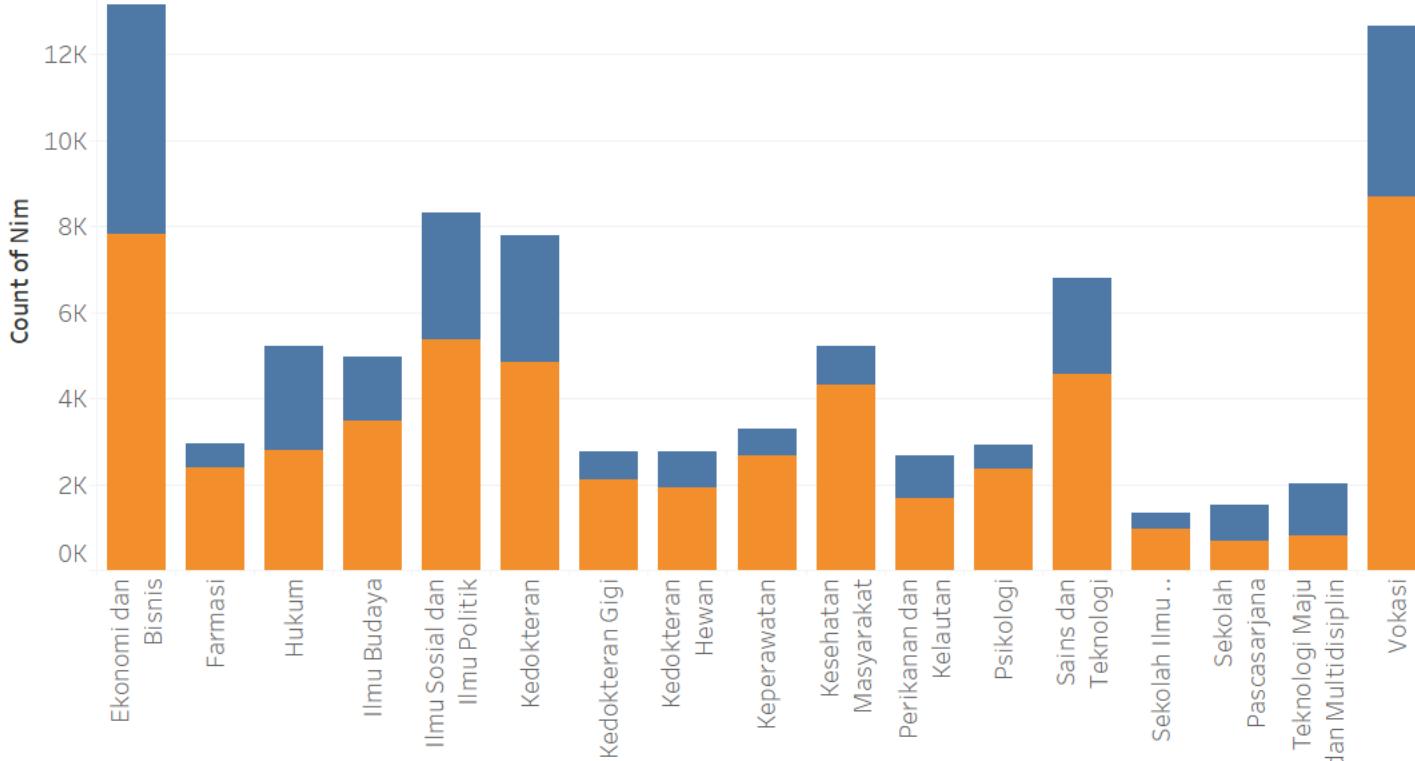


BETTER ALTERNATIVE

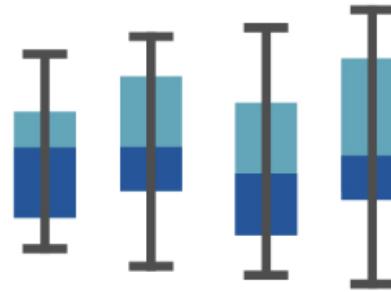




Lebih efektif yang mana?

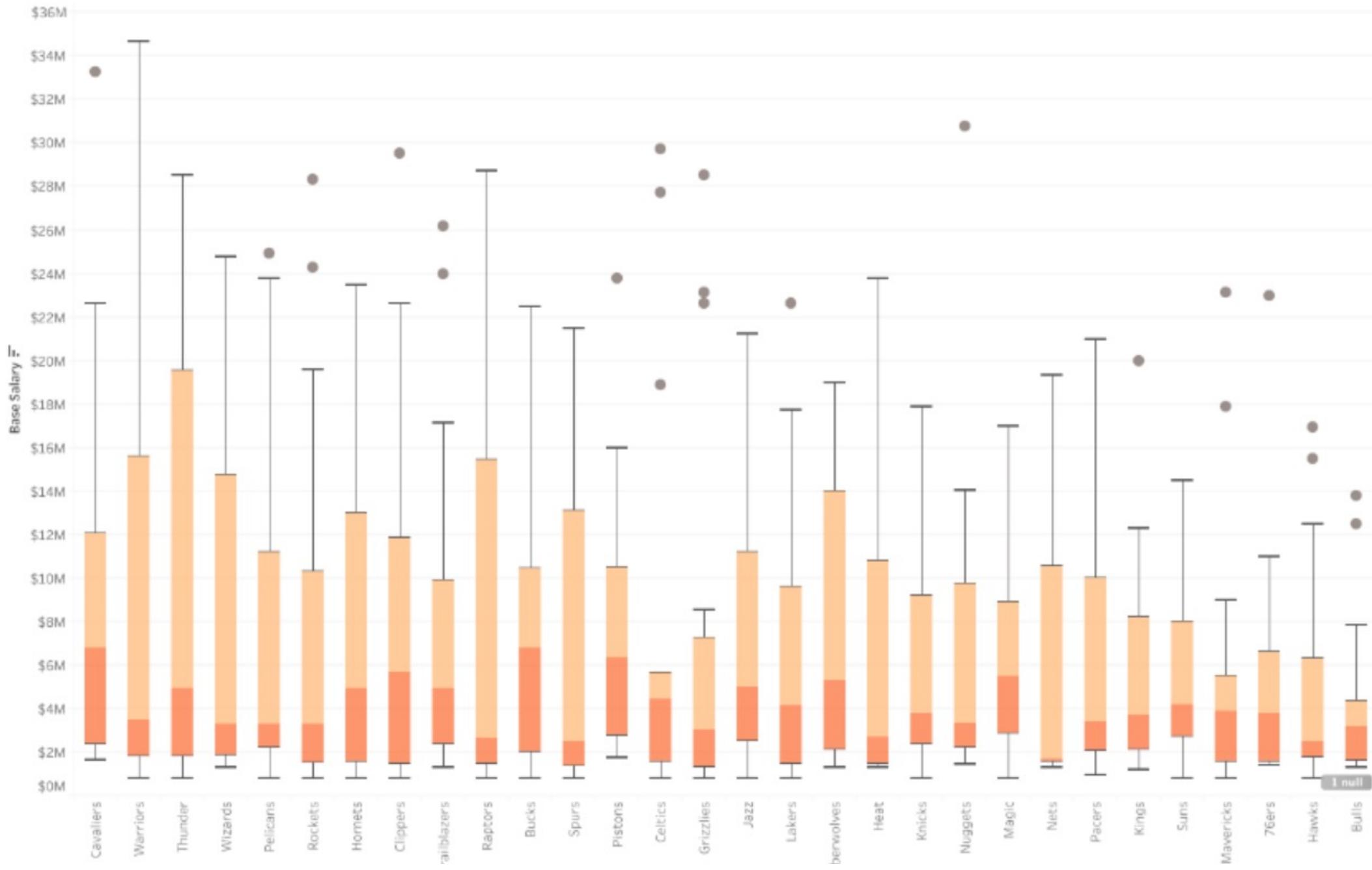


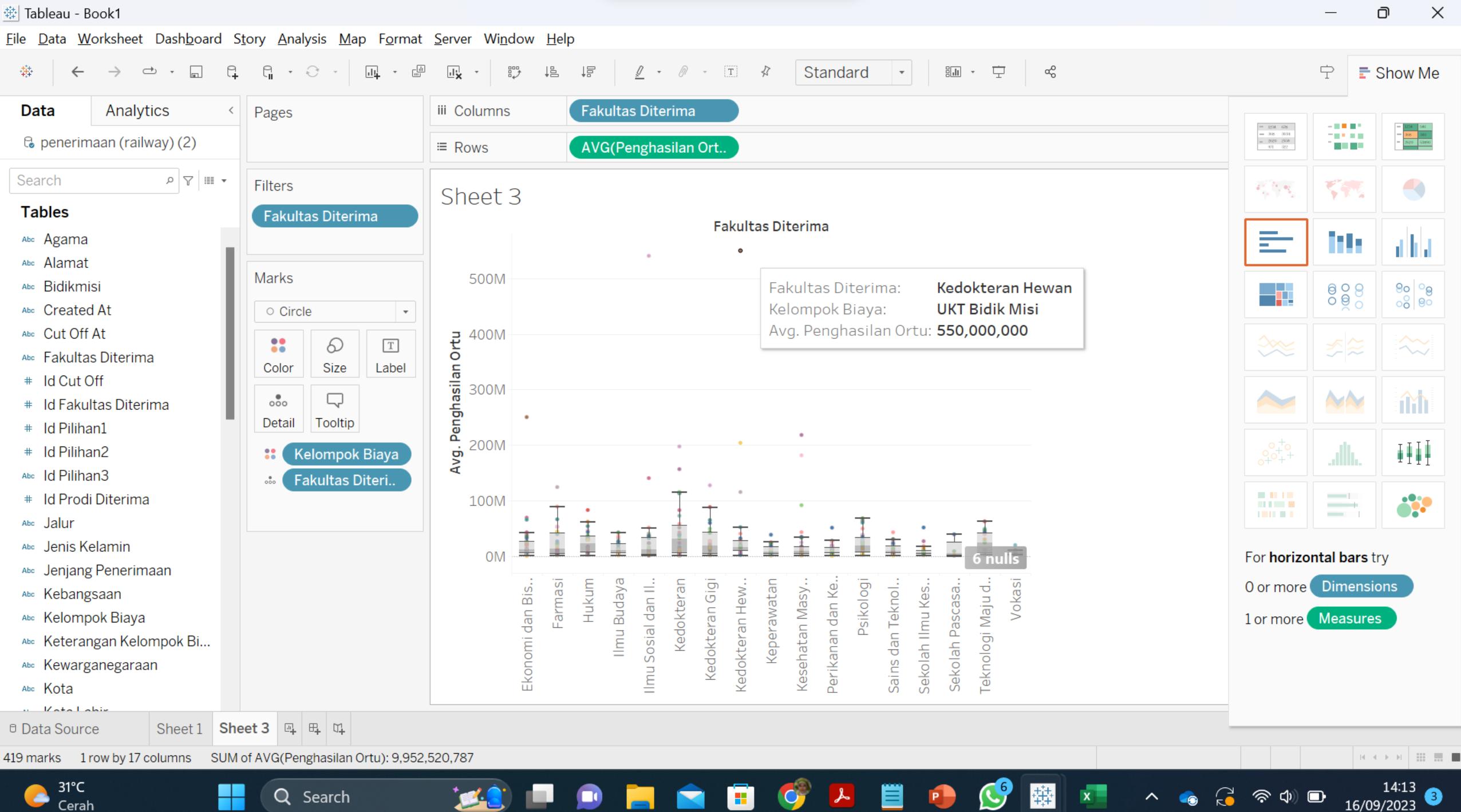
Box and Whisker Plot

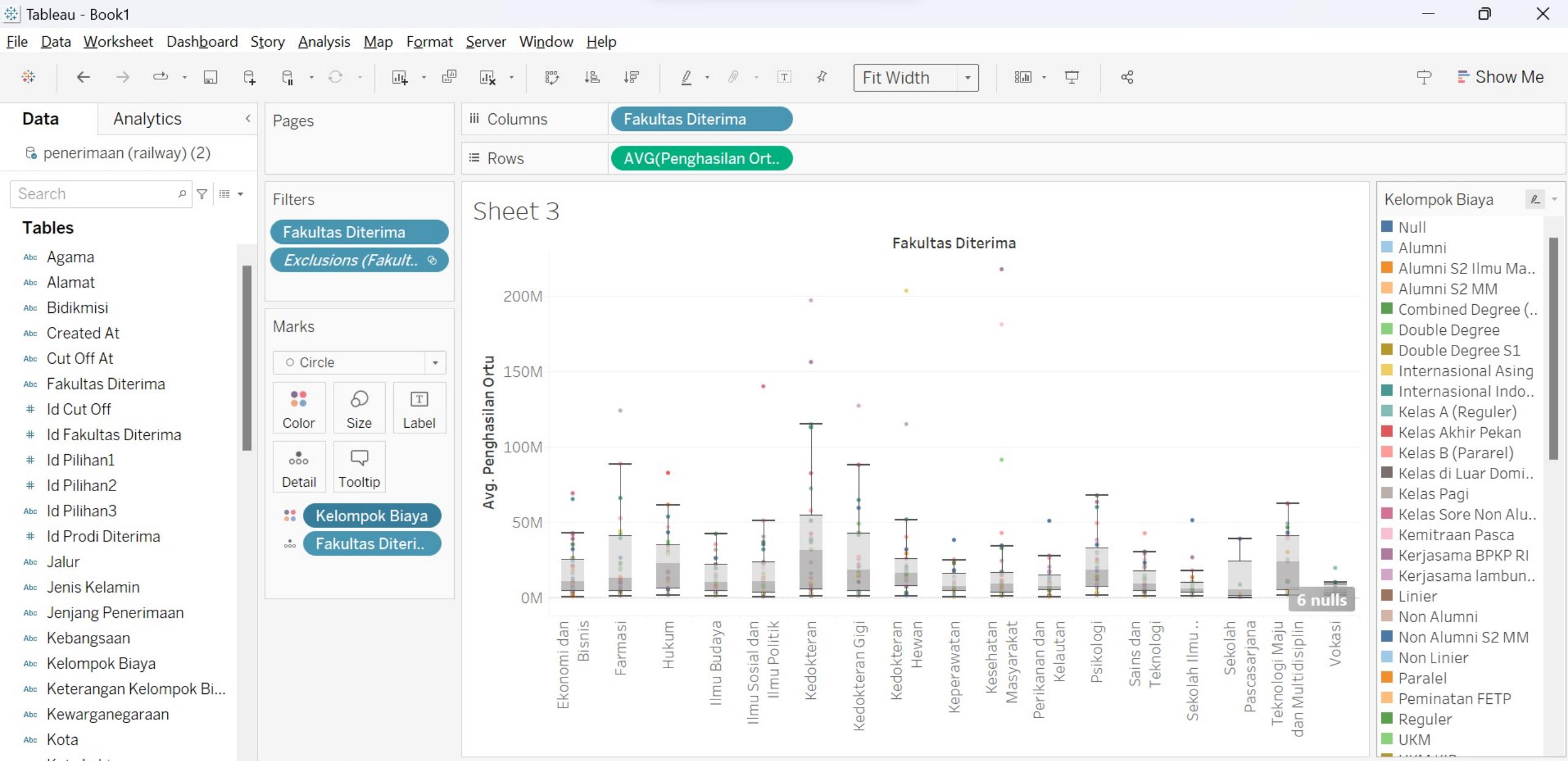


- Box and whisker plots portray the distribution of your data, outliers, and the median. It summarizes a data set in five marks. The mark with the greatest value is called the maximum. The mark with the lowest value is called the minimum.
- The box itself contains the lower quartile, the upper quartile, and the median in the center. The median is the value separating the higher half from the lower half of a data sample, a population, or a probability distribution.
- The whiskers (the lines extending from the box on both sides) typically extend to 1.5^* the Interquartile Range (the box) to set a boundary beyond which would be considered outliers. Hence the name, box, and whisker plot.

Which NBA team had the widest range of salaries?
Teams sorted descending by total salary. Gray dots are outliers







Data Source Sheet 1 Sheet 3

416 marks 1 row by 17 columns SUM of AVG(Penghasilan Ortu): 8,611,839,281

31°C
Mendekati rekord



Search



14:14
16/09/2023

Gantt Chart

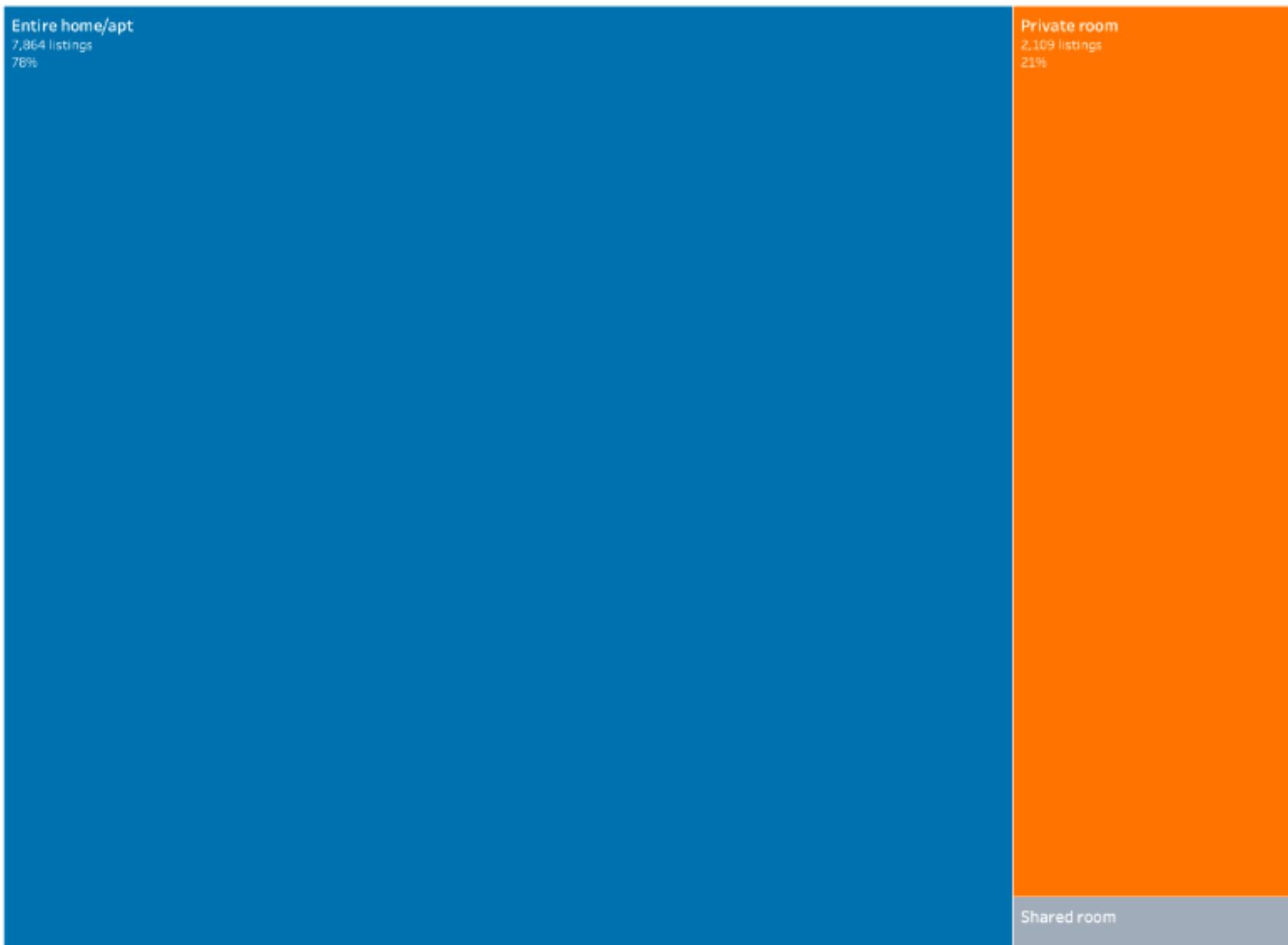
- Use Gantt charts to show the duration of events or activities.
- In a Gantt chart, each separate mark (usually a bar) shows a duration. For example, you might use a Gantt chart to display average delivery time for a range of products.
- Gantt charts can also used for monitoring load for each worker.



Treemap

- The treemap functions as a visualization composed of nested rectangles.
- These rectangles represent certain categories within a selected dimension and are ordered in a hierarchy, or “tree.”
- Quantities and patterns can be compared and displayed in a limited chart space.

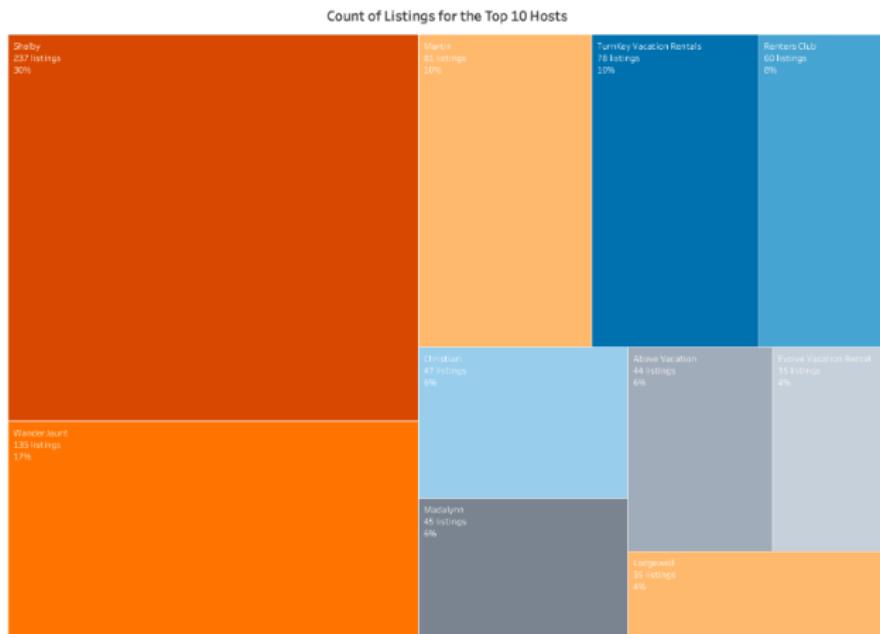
Listings per Room Type



POOR EXAMPLE

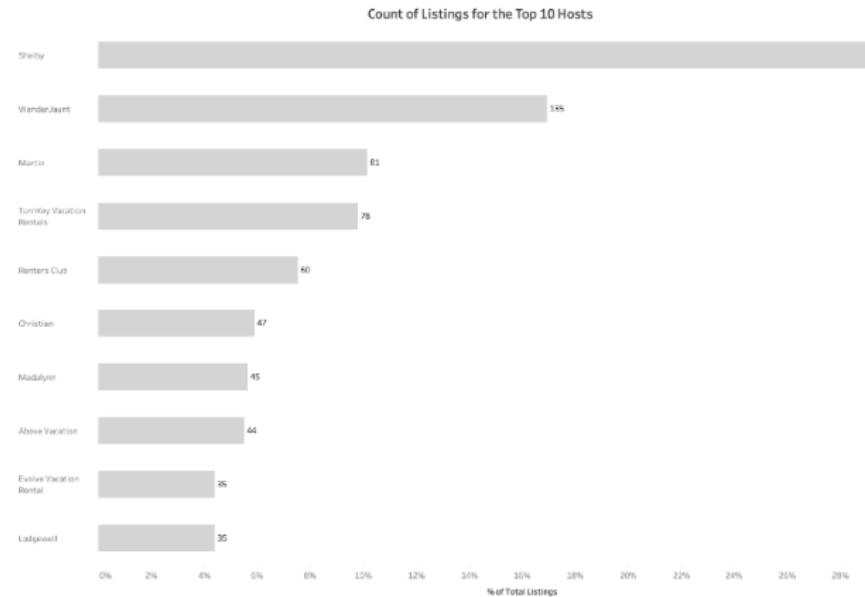
This treemap looks at the top ten hosts in a region for Airbnb and measures the categories by a number of listings. Unfortunately, the map is confusing and difficult to read.

- The same color is used for multiple categories
- Many of the boxes look similar in size and measure



BETTER ALTERNATIVE

A better alternative for an ineffective treemap would be a simple [bar chart](#). This chart can be organized from largest to smallest, so the reader can easily compare to see who has the most listings and who has the fewest listings.

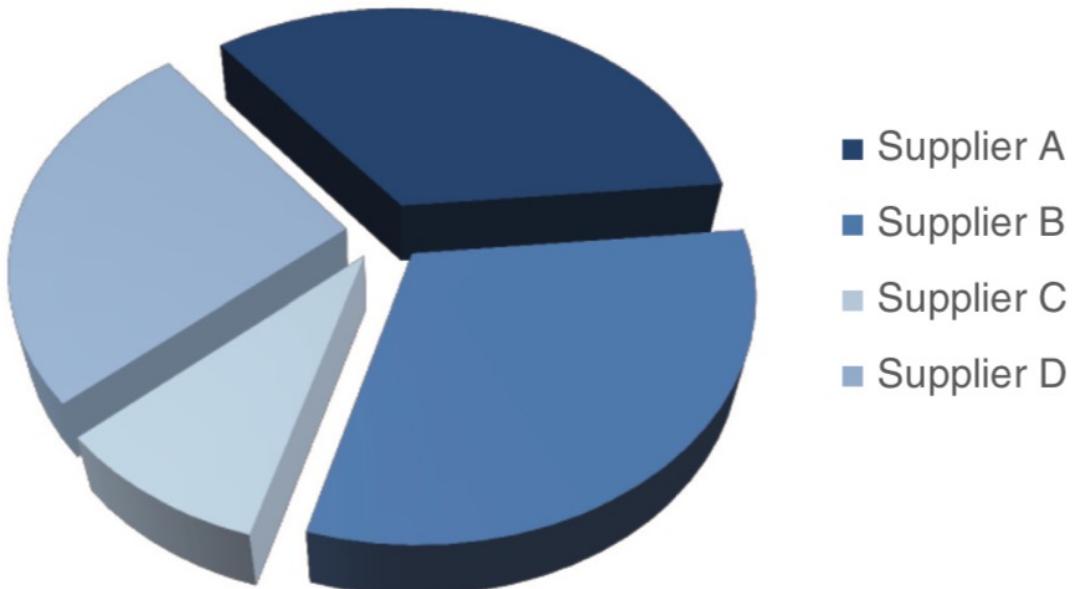


TO BE
AVOIDED

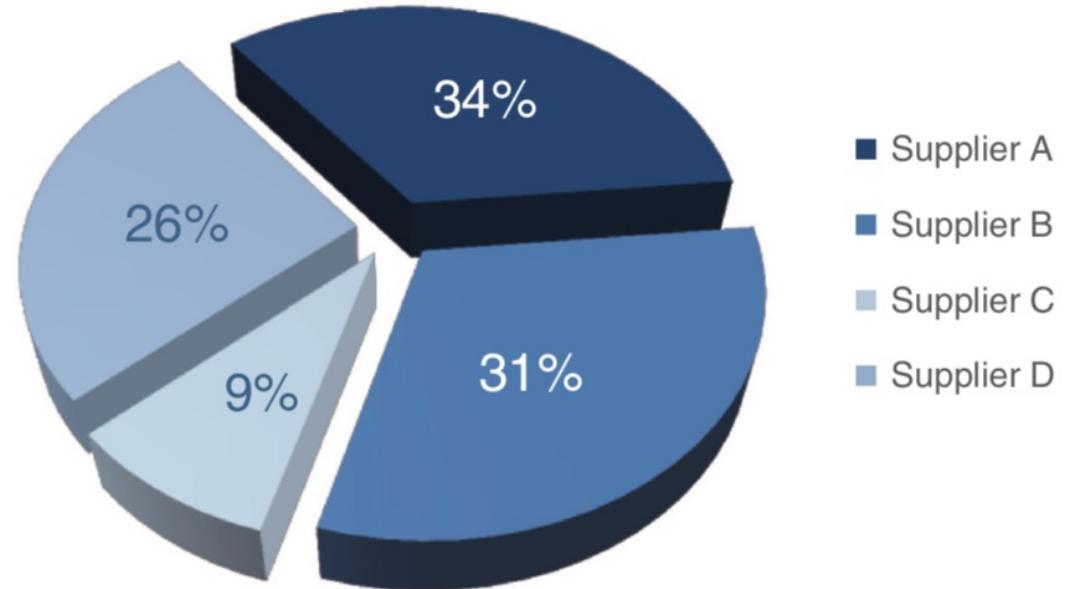


Pie charts

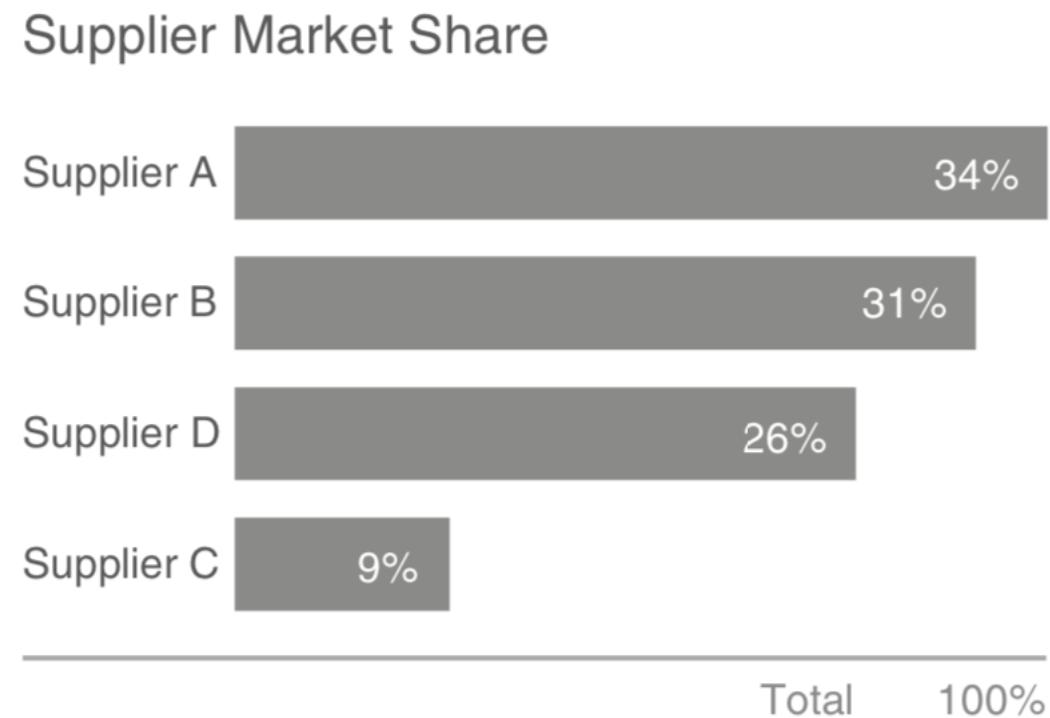
Supplier Market Share



Supplier Market Share

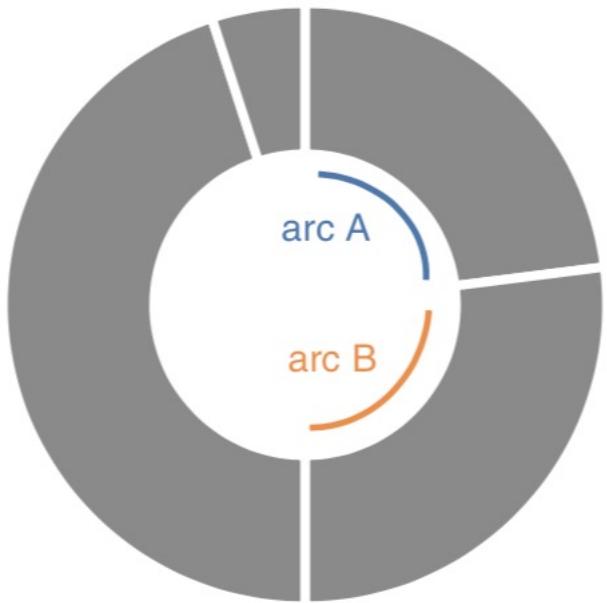


Alternative to pie chart

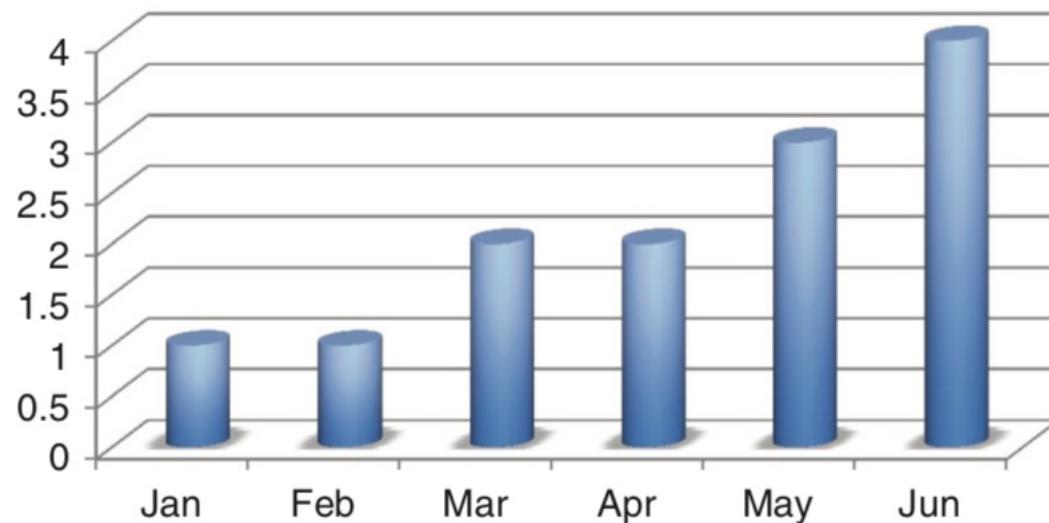


Donut and 3D chart

The donut chart



Number of issues



Secondary y-axis

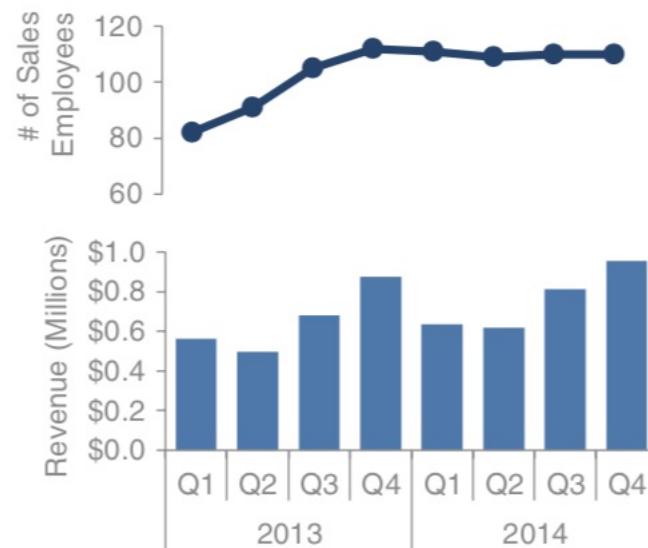


Alternative to secondary y-axis

Alternative 1: label directly



Alternative 2: pull apart vertically



References

Knafllic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Knafllic, C. N. (2019). *Storytelling with Data: Let's Practice!*. John Wiley & Sons.

<https://www.storytellingwithdata.com/>

<https://www.youtube.com/@storytellingwithdata>