Data Visualization 101

Perhaps, you might have asked yourself..

What is Data Visualization?

What should I consider before creating a visualization?

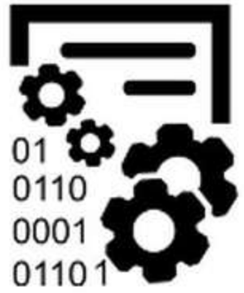
Which visualization should I use?

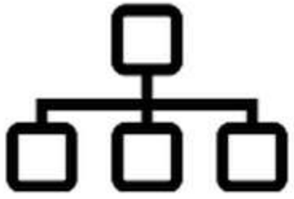
What are the rules of good visual design?

etc.

Your
Data Visualization
journey starts here!

3-4-5 Visualization Framework





3
Questions to ask yourself

4
Types of visualizations

5
Rules of good visual design

Before we start..

What is Data Visualization?

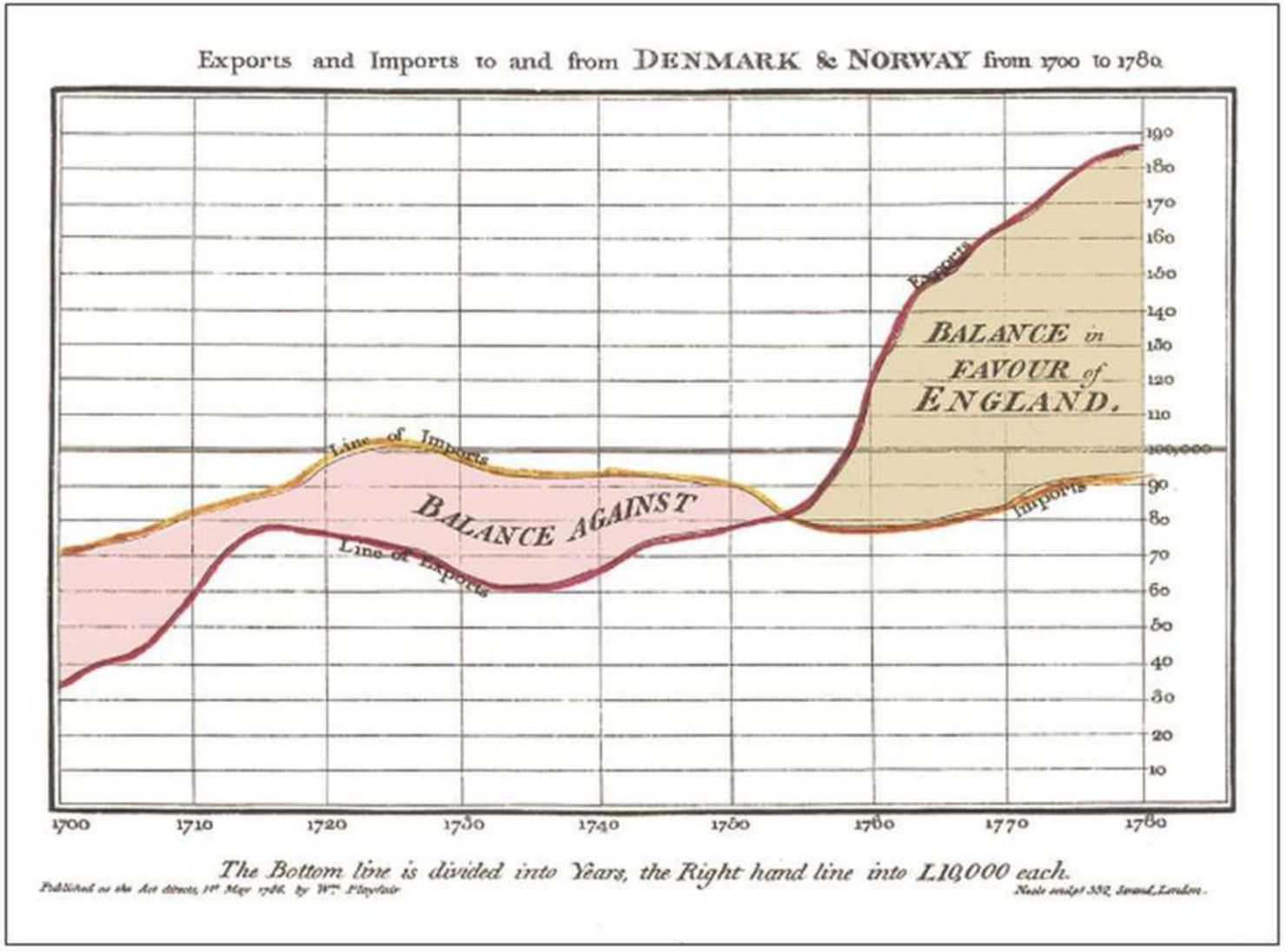
Data Visualization…

…is a form of visual **communication**.

…aims to convey information **clearly** and **efficiently**.

…makes complex data more **accessible** and **understandable**.

…is both an **art** and **science**.



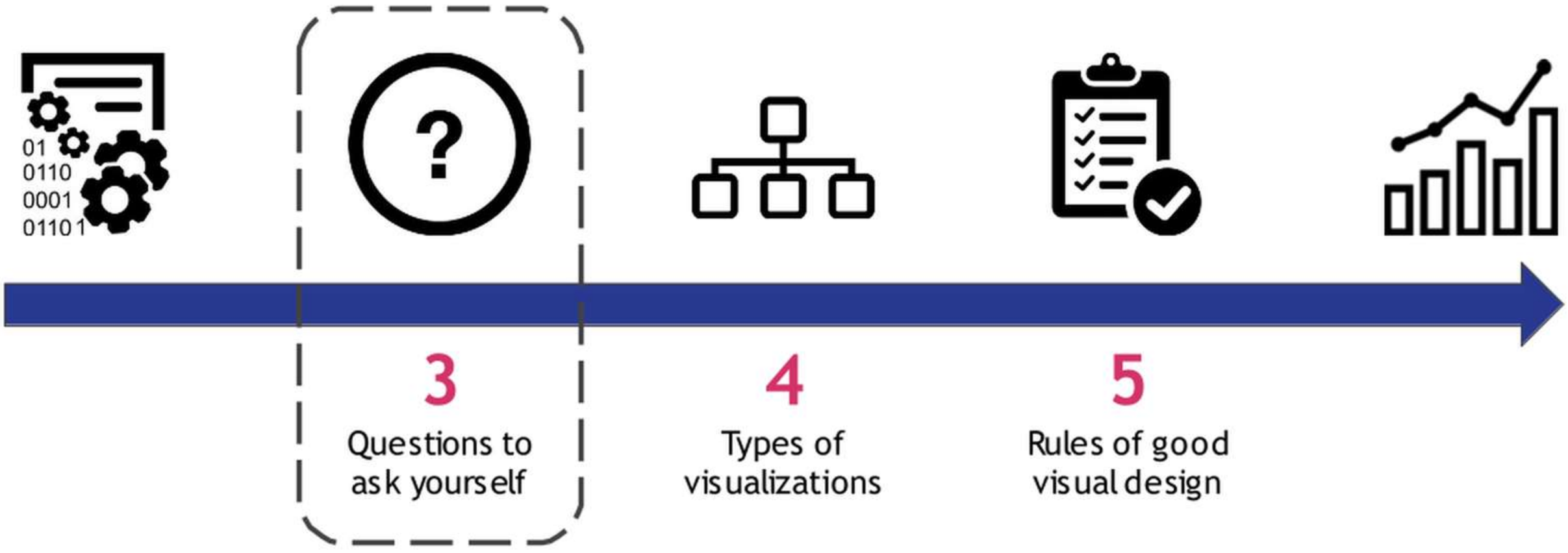
Visualization is not a modern invention...

Source: Playfair's Commercial and Political Atlas

...but it is becoming **increasingly important** in today's world.

Source: Facebook

3-4-5 Visualization Framework



3 Questions

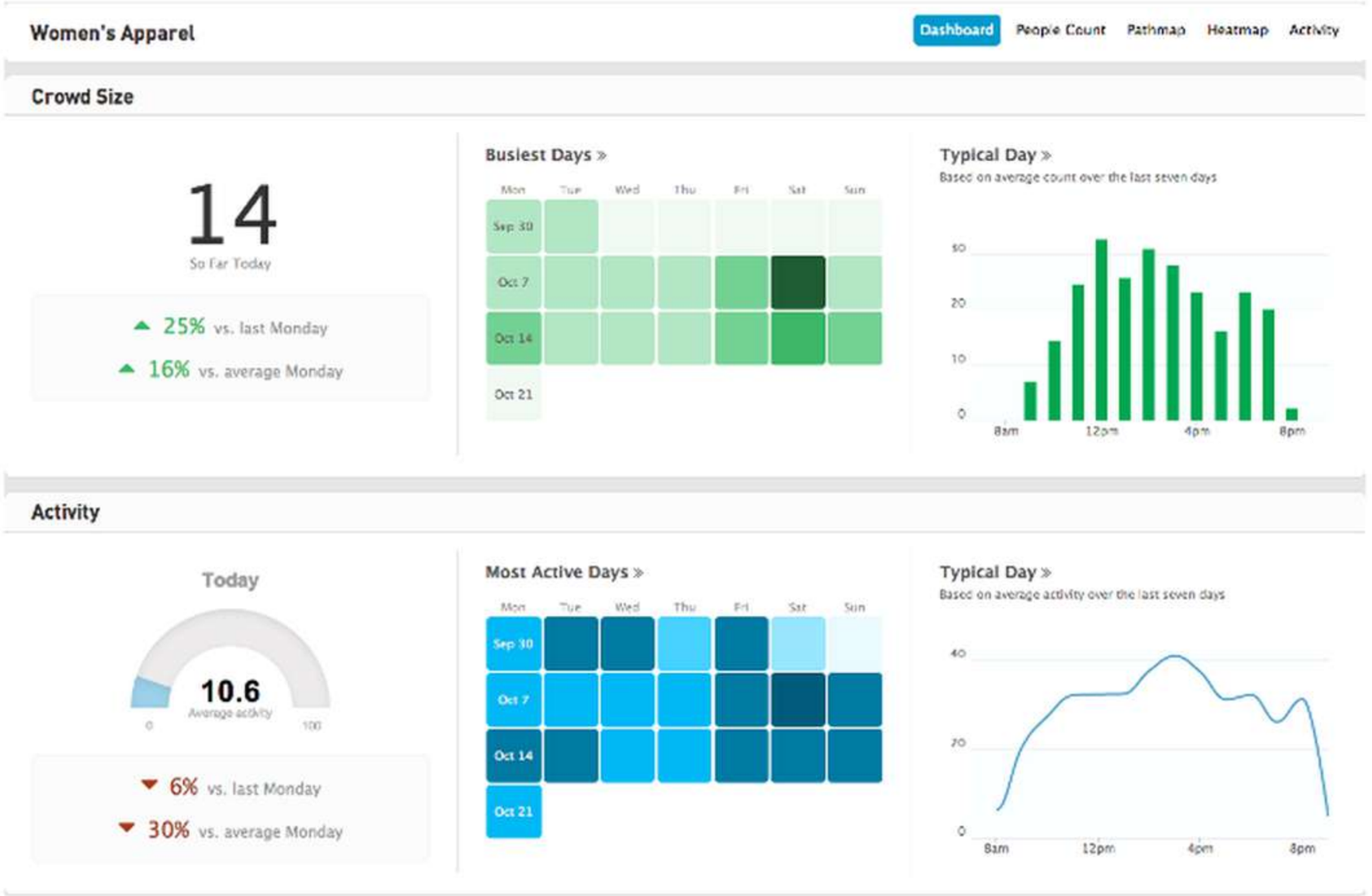
to ask yourself at the **start.**

1. Who is my **audience**?
2. What is my **message**?
3. Is it **worth creating** a visualization?

1. Who is my **audience**?

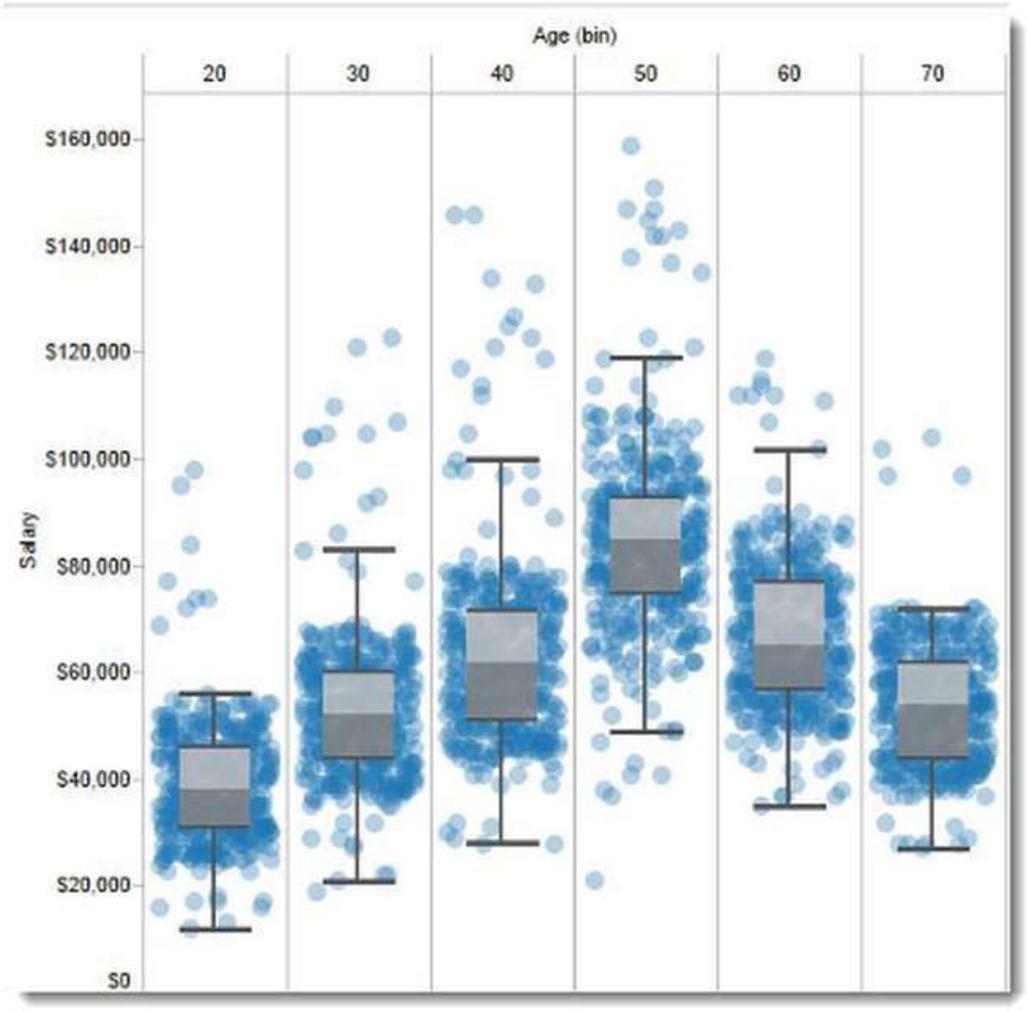
Understand the **background** and **expectations** of the visualization audience.

Background	**Experts**	vs.	**Managerial**
Expectations	**Quick glance**	vs.	**Deep-dive**
Guidance	**Instructed**	vs.	**Self-discovery**



An example of a **managerial self-service** dashboard.
Its visualizations are quickly and easily understood.

Source: HBR



An example of a more complex **Box Plot** visualization.
Usually for **expert** audiences or requires **guided** interpretation.

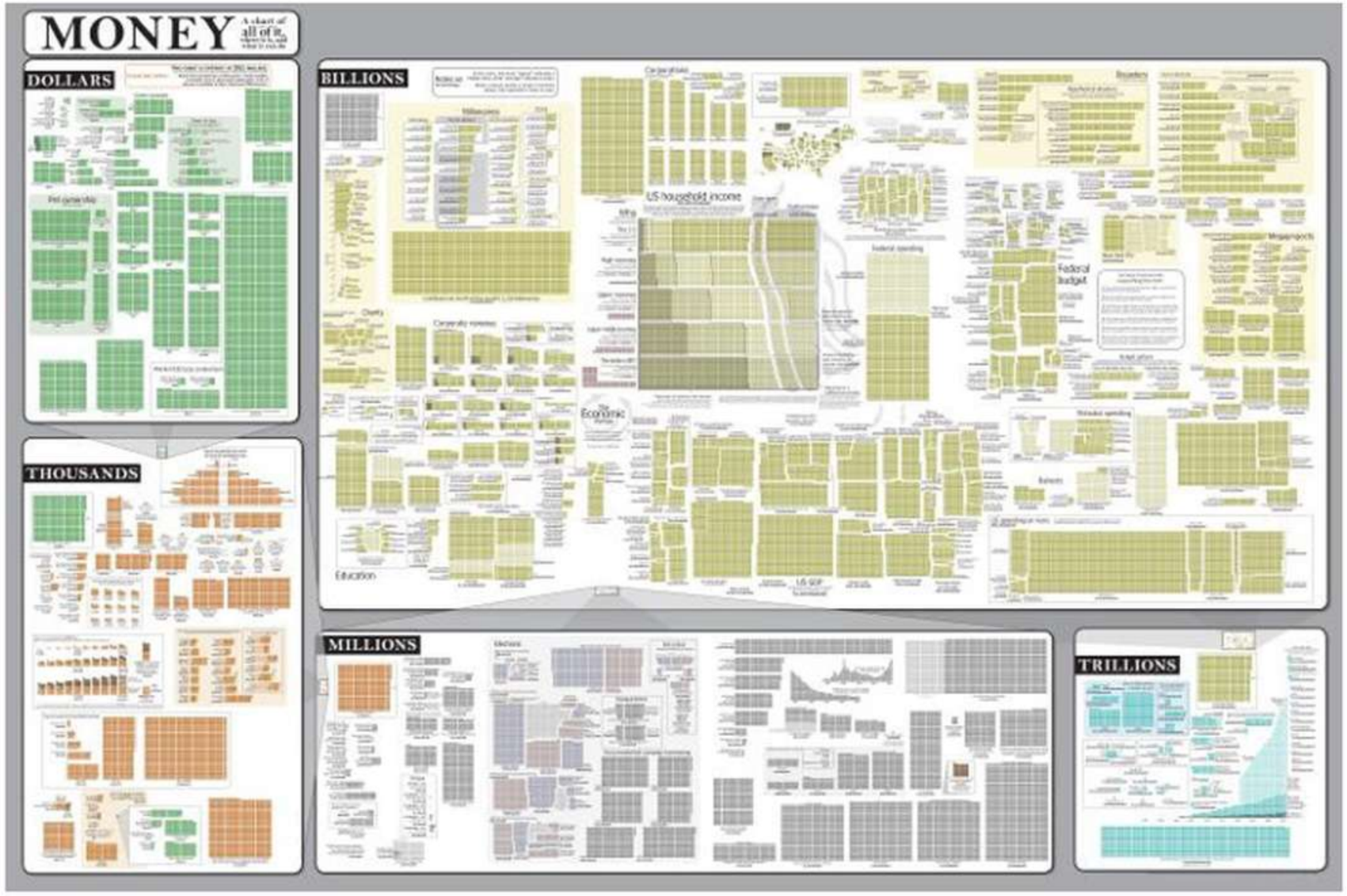
Source: Data Revelations

2. What is my **message**?

Data **by itself** does not tell a story.

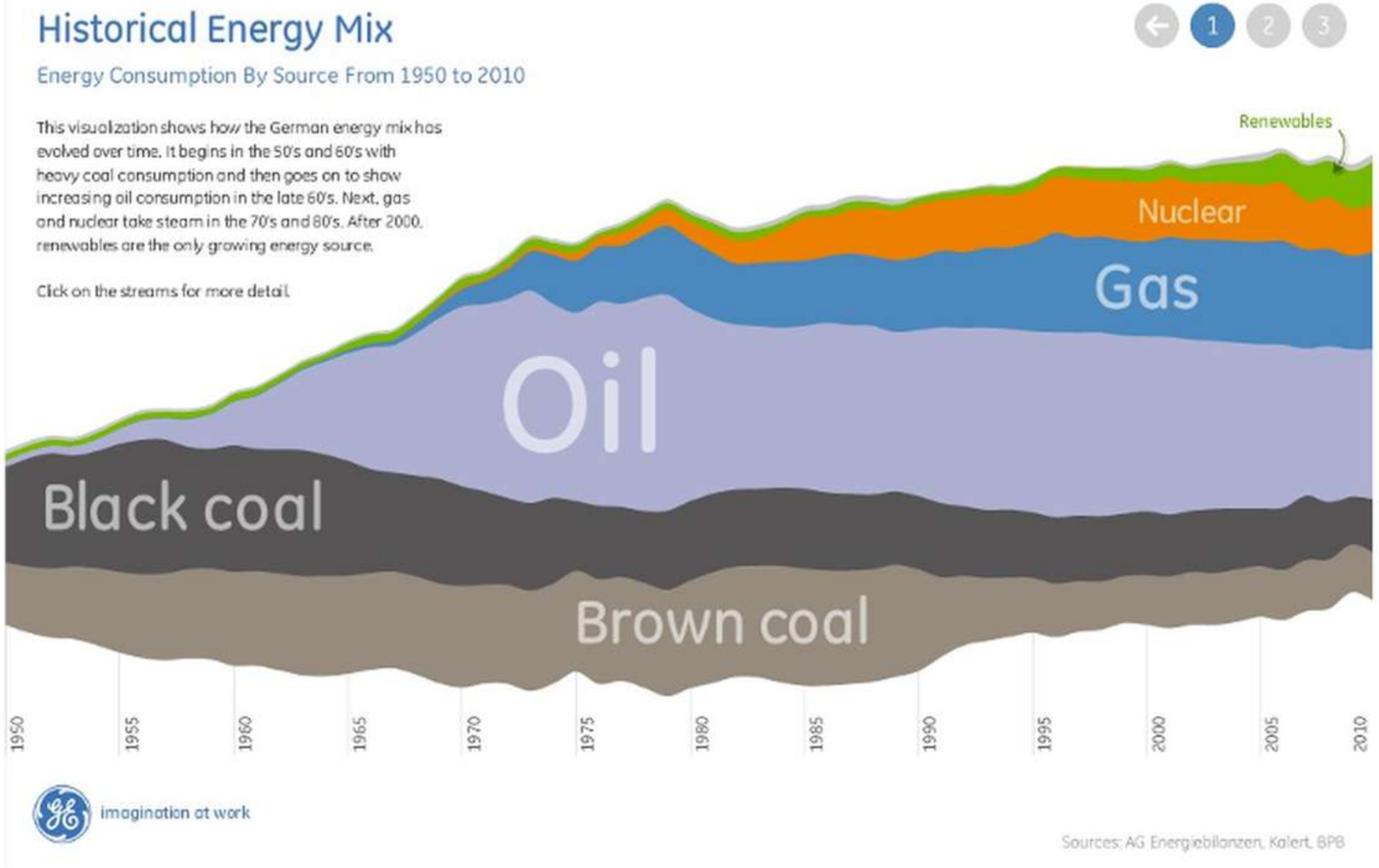
What **key idea** do you want your audience to take away?

Craft your **message track** before creating your visualization.



Data by itself does not tell a story. It is just information.

Source: xkcd



A example of a good visualization with a clear story.

Source: GE

3. Is it **worth creating** a visualization?

Does your visualization provide an insight that was **not obtainable** with the original representation of data?

Visualizations are not a cure-all - sometimes a table might work better!

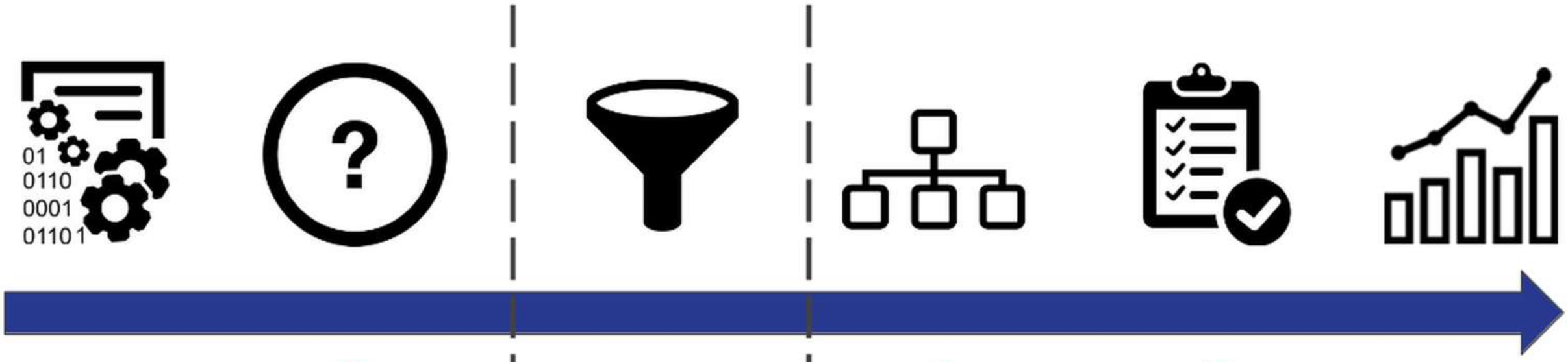
Tables	Visualizations
Data as text	Data as pictures
Data arranged in **rows & columns**	Data displayed in relation to **axes**
Precise, individual values	Message resides in the data **shape**

Recap

3 questions to ask yourself.

1. Who is my **audience**?
2. What is my **message**?
3. Is it **worth creating** a visualization?

3-4-5 Visualization Framework



3
Questions to ask yourself

Data Preparation

4
Types of visualizations

5
Rules of good visual design

Data Preparation

Raw data is **rarely perfect**.

There might be missing values and incorrect data types.

Depending on your **message**, you may also need to filter out or create new data.

Examine your data and prepare it via:

1. **Cleaning:** To handle noise, missing values, NULL values, etc.

2. **Transformation:** Aggregating, filtering, new calculated fields, pivoting, etc.

Country	GDP per capita
Qatar	N/A
Luxembourg	91
Norway	65
Switzerland	54
United States	53
Germany	43
United Kingdom	36
Romania	19
China	N/A
Albania	N/A

Country	GDP per capita
Luxembourg	91
Norway	65
Switzerland	54
United States	53
Germany	43
United Kingdom	36
Romania	19

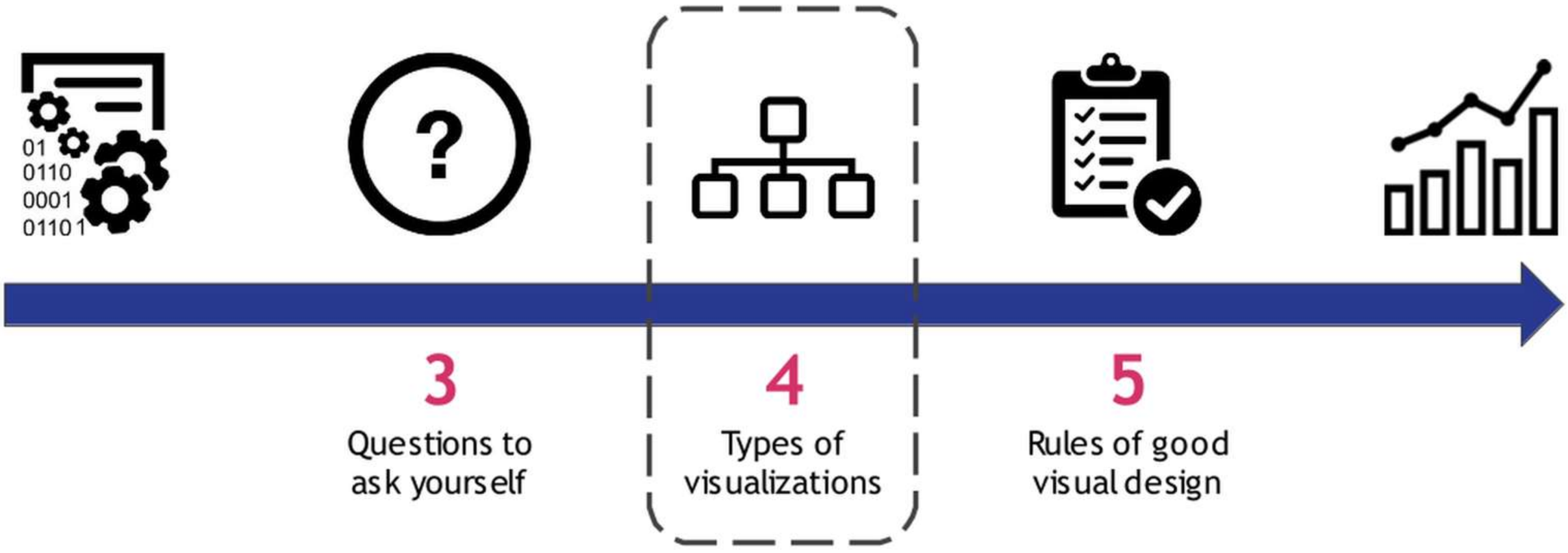
An example of data cleaning to handle missing values.

Music	Labels
Pop	87
Dance	76
House	12
Rock	62
Punk	54
Classical	53
Soul	43
R&B	36

Group	Labels
Pop, Dance, House	175
Rock & Punk	116
Classical	53
Soul & R&B	79

An example of data transformation by aggregation.

3-4-5 Visualization Framework



What would you like to show?

Visualizations can be categorized according to their **intent.**

There are **4** main visualization types:

1. Comparison
2. Distribution
3. Relationship
4. Composition

4 Visualization Types: **Comparison**

Intent is to compare either **between items**, **over time** or **both**.

What are examples of each?

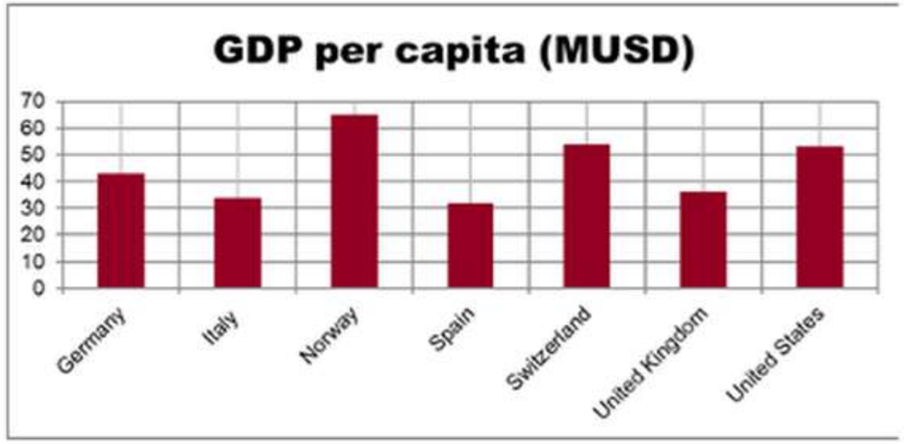
Between items: # of labels by music genre, Sales for a CD…

Over time: Monthly average office temperature, Quarterly sales…

Both: Weekly usage hours by application, Monthly sales by CD…

4 Visualization Types: **Comparison**

Between items: Comparing GDP per capita by country.



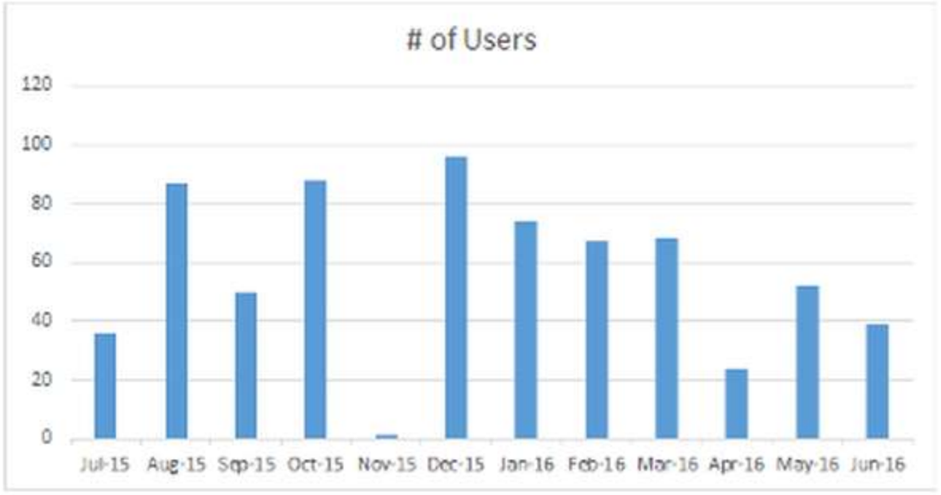
What is the difference?

Horizontal labels and bars for ease of reading.

Quantitative values are **sorted** in order.

4 Visualization Types: **Comparison**

Over time: Monthly # of users for FY15/16

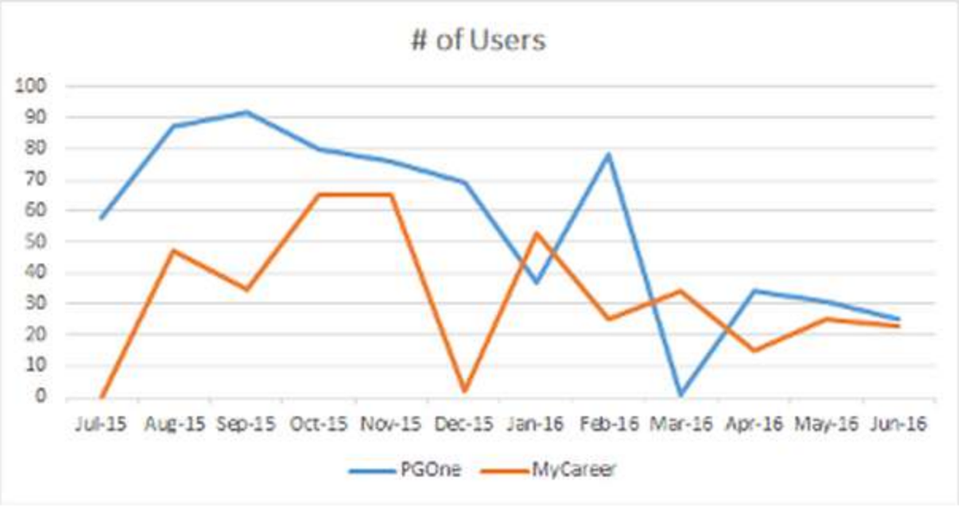


Time dimension on the **horizontal** axis helps indicate the flow of time (from left to right).

Height of Bars illustrate relative magnitude difference across months.

4 Visualization Types: **Comparison**

Both among items and over time: Monthly # of users for FY15/16 by application.



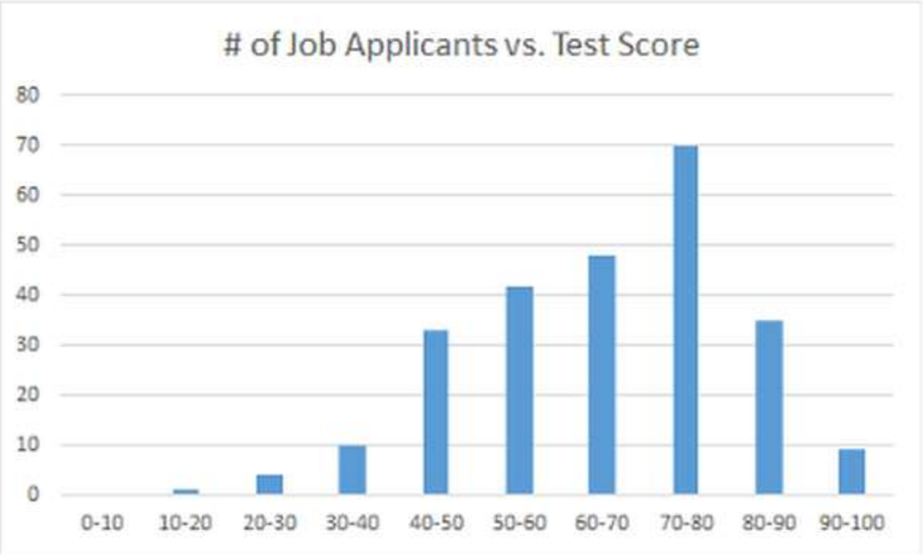
A **line chart** connects the same item across time periods and helps illustrate **trend**.
Multiple lines helps to compare between different items at individual time periods.

4 Visualization Types: **Distribution**

Intent is to illustrate the **spread** of data, possibly across defined groups.

What are some examples?

Applicant test scores, # of users across 24 hours, etc.



A **histogram** depicting the spread of test scores by applicants.
Is the test too easy or difficult?

4 Visualization Types: **Relationship**

Intent is to show the **relationship** between **two or more variables.**

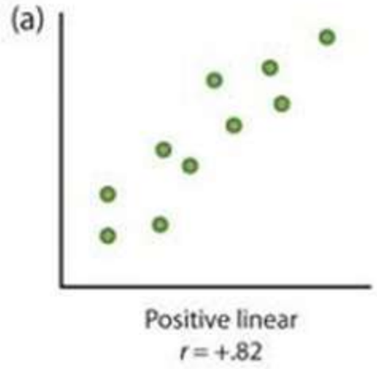
What are some examples?

CD sales vs. price, # of hours in office vs. salary, etc.

A **scatter plot** illustrates the relationship between 2 variables. Is this a positive or negative relationship?

4 Visualization Types: **Relationship**

Scatter Plots are useful to see the relationships between variables.

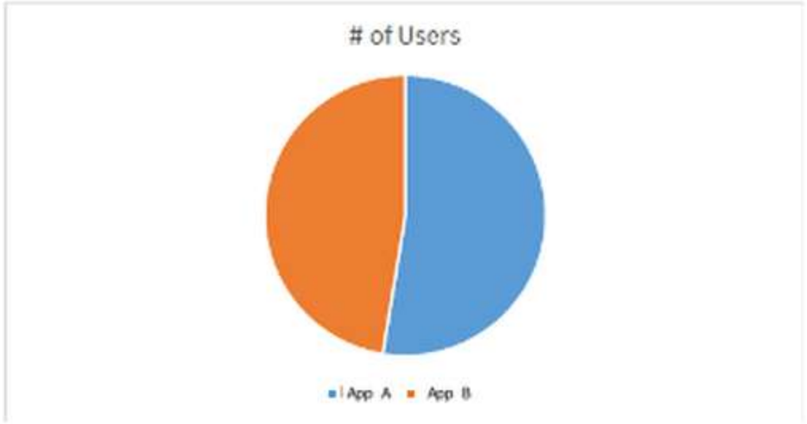


4 Visualization Types: **Composition**

Intent is to see individual data segments **as part of a whole.**

Static

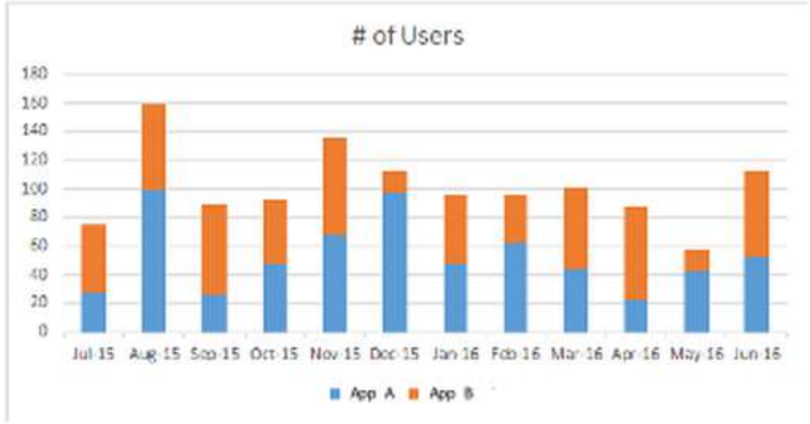
E.g., User breakdown by application



Pie Chart

Changing over time

E.g., Monthly User breakdown by application



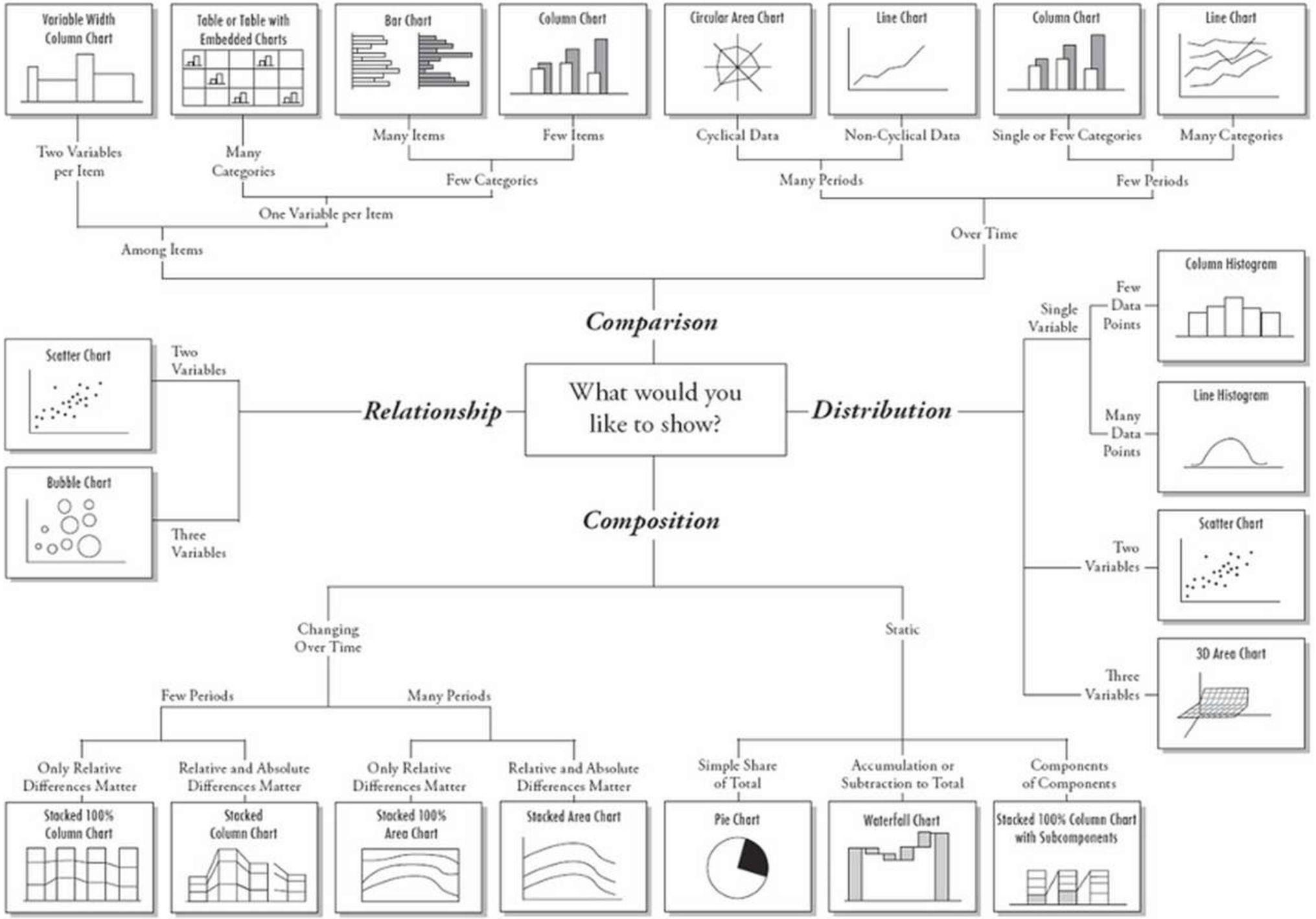
Stacked Bar Chart

Recap

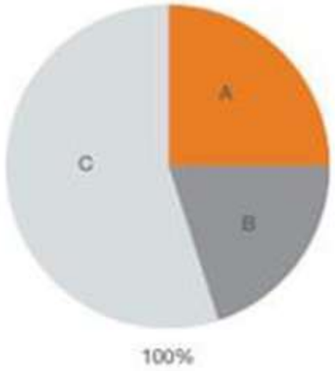
4 types of visualizations.

1. Comparison
2. Distribution
3. Relationship
4. Composition

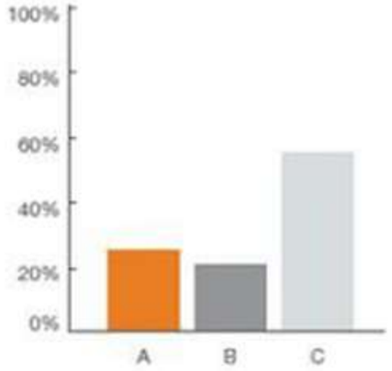
Visualization Cheat Sheet



Visualization Tips: Pie vs. Bar Charts



Pie charts work only for showing large differences in proportion, especially percentages. Use them when you want to show all of the parts that make up a whole, or compare the percentages of one set to the percentages of another.



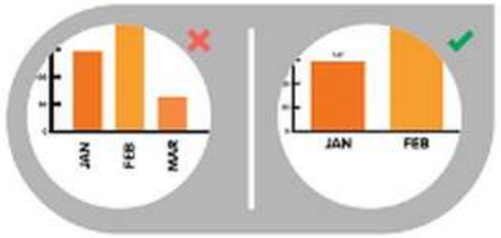
Bar charts are visually more precise than pie charts, and can accommodate larger data sets. Plus, you can stack them to add an additional set of data. Use them when you need to show precise relationships.

Source: Slide:ology – Nancy Duarte

Visualization Tips: Bar Charts

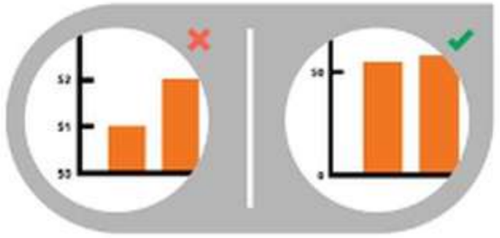
BAR CHART

DESIGN BEST PRACTICES



USE HORIZONTAL LABELS
Avoid steep diagonal or vertical type, as it can be difficult to read.

SPACE BARS APPROPRIATELY
Space between bars should be ½ bar width.



START THE Y-AXIS VALUE AT 0
Starting at a value above zero truncates the bars and doesn't accurately reflect the full value.

USE CONSISTENT COLORS
Use one color for bar charts. You may use an accent color to highlight a significant data point.

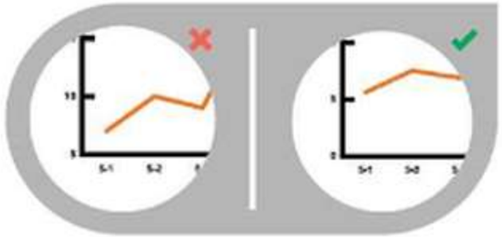
ORDER DATA APPROPRIATELY
Order categories alphabetically, sequentially, or by value.

Source: Visage and Hubspot "Data Visualization 101", via Slideshare

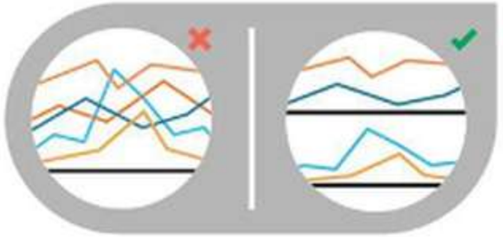
Visualization Tips: Line Charts

LINE CHART

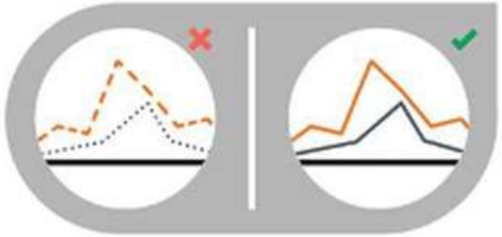
DESIGN BEST PRACTICES



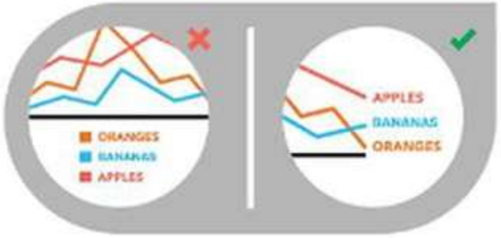
INCLUDE A ZERO BASELINE IF POSSIBLE
Although a line chart does not have to start at a zero baseline, it should be included if possible. If relatively small fluctuations in data are meaningful (e.g., in stock market data), you may truncate the scale to showcase these variances.



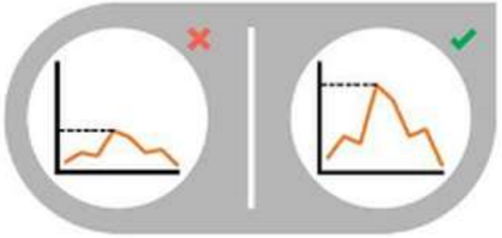
DON'T PLOT MORE THAN 4 LINES
If you need to display more, break them out into separate charts for better comparison.



USE SOLID LINES ONLY
Dashed and dotted lines can be distracting.



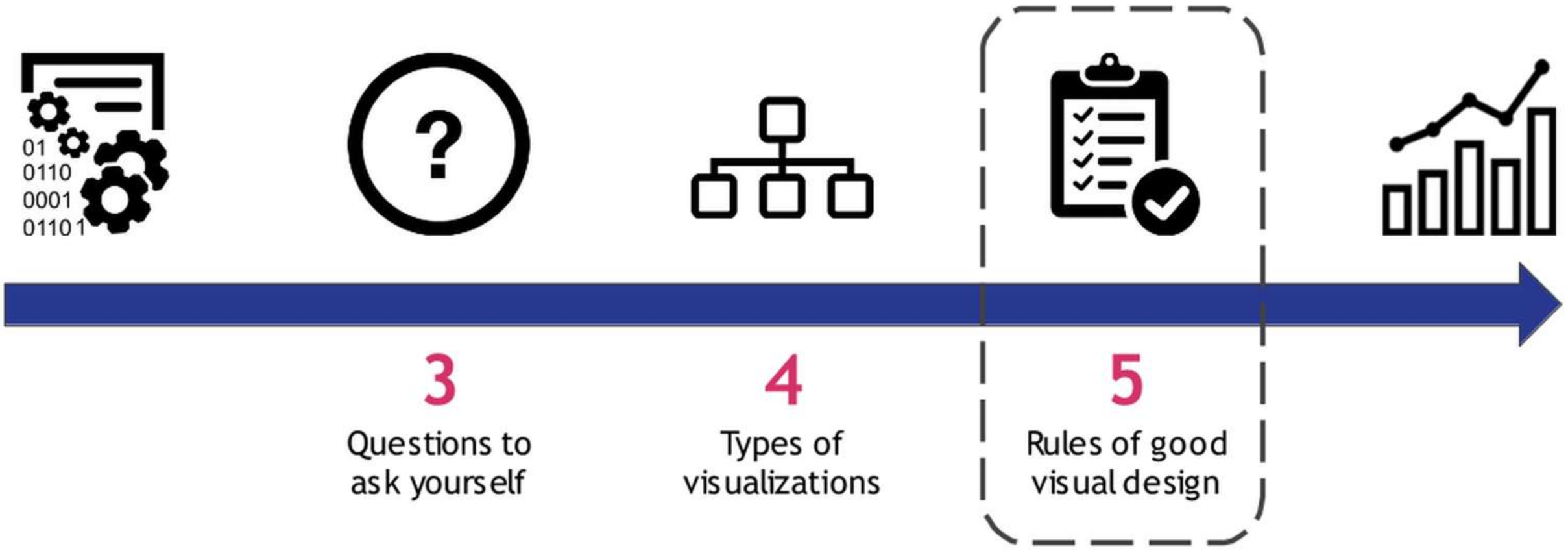
LABEL THE LINES DIRECTLY
This lets readers quickly identify lines and corresponding labels instead of referencing a legend.



USE THE RIGHT HEIGHT
Plot all data points so that the line chart takes up approximately two-thirds of the y-axis' total scale.

Source: Visage and Hubspot "Data Visualization 101", via Slideshare

3-4-5 Visualization Framework



5 Rules of Good Visual Design

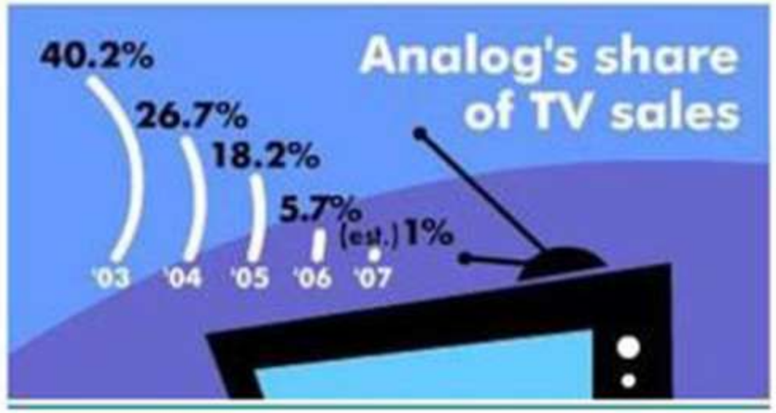
1. No **noise** (a.k.a. Chartjunk).
2. Use **colours** wisely.
3. Avoid using **3D** effects.
4. No misleading **scales**.
5. Be careful with **dual-axis** charts.

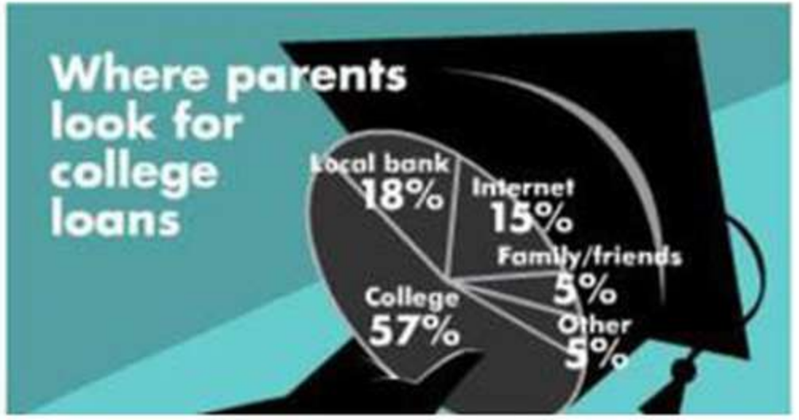
1. No **noise** (a.k.a. Chart junk)

Chart Junk:

Visual content that adds **no value** and **distracts** from the data.

Examples: Shadows, colour gradients, graphics, etc.





Are gridlines chart junk?

Gridlines are **often** chart junk.

Use them (with care) to:
- Increase precision on **wide** graphs.
- Emphasize small differences.

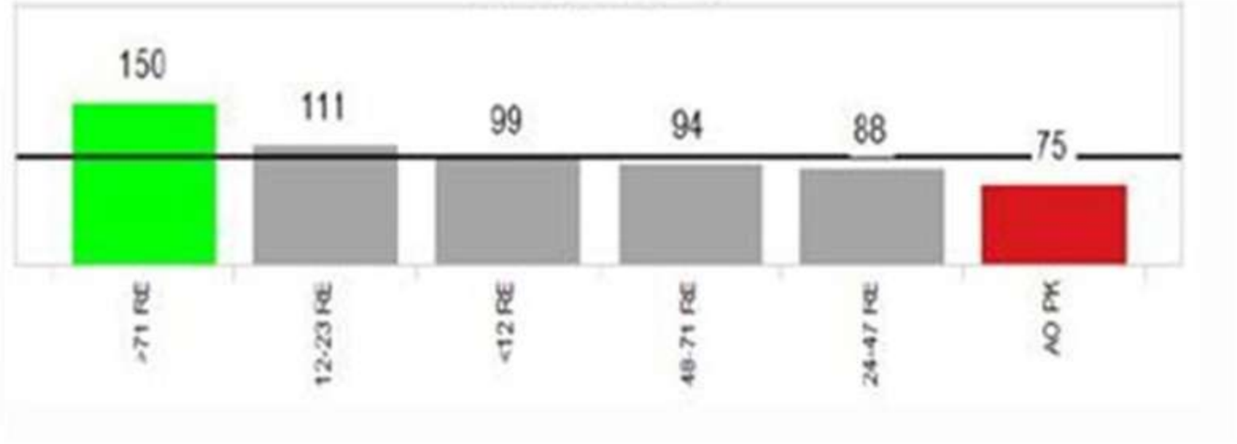
If used, they should be visually **muted**.

2. Use **colours** wisely

Intense colours only to draw attention.

Different colours to indicate differences in data.

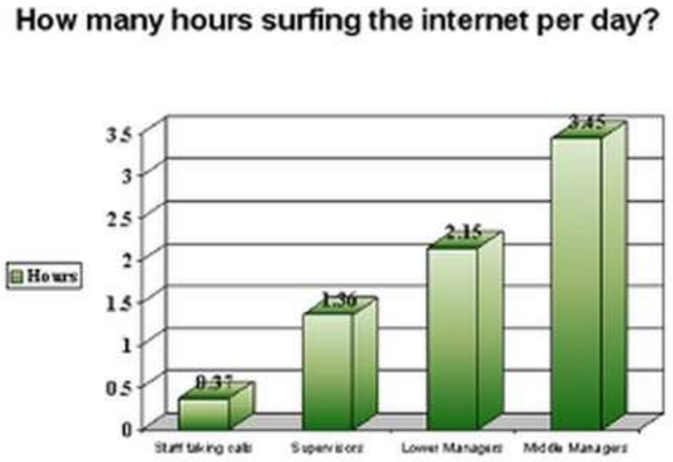
Single, neutral **background** colour (if needed at all).



3. Avoid using **3D** effects

3D effects are a form of **chart junk.**

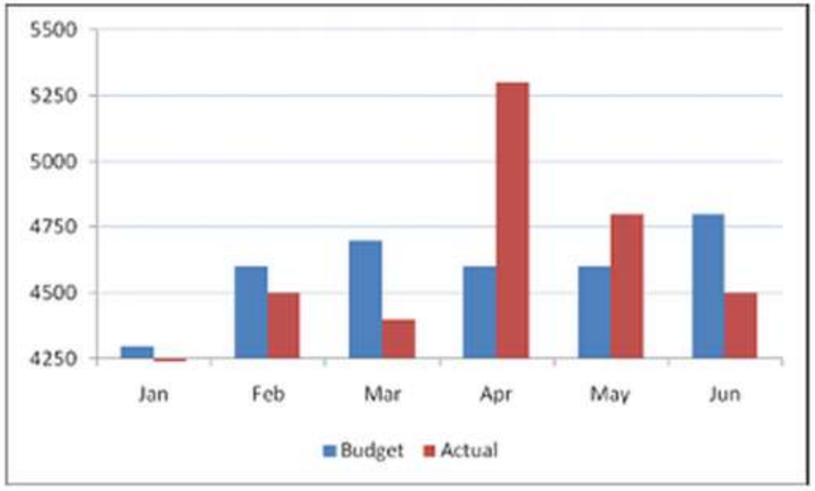
Deserves special mention because of its pervasiveness and how it **gets in the way of communication.**



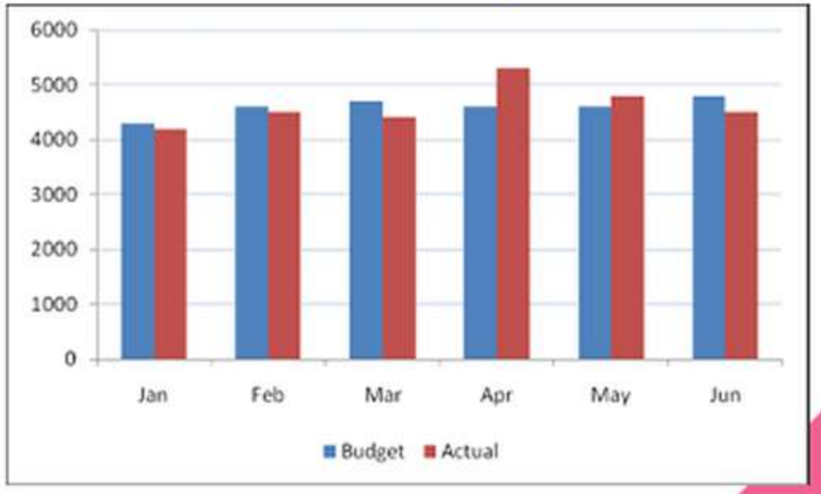
4. No misleading **scales**

If your axes **do not start at 0**, it might mislead audience perception, especially for **comparison** visualizations. Send the correct message!

Our April expenditure is >2x our budget!! What is Finance doing?!



Oh...



5. Be careful with **dual-axis** charts

Dual-axis charts are only useful when comparing data with **different units of measure**.

Even then, **2 separate charts** might be more effective.

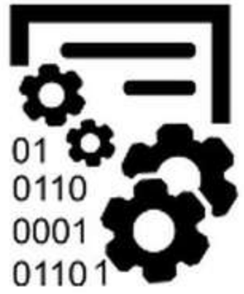
What is this chart telling me??

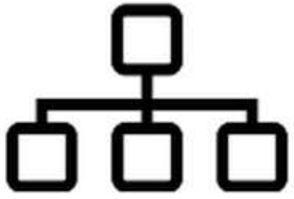
Recap

5 rules of good visual design.

1. No **noise** (a.k.a. Chart junk).
2. Use **colours** wisely.
3. Avoid using **3D** effects.
4. No misleading **scales**.
5. Be careful with **dual-axis** charts.

3-4-5 Visualization Framework





3
Questions to ask yourself

4
Types of visualizations

5
Rules of good visual design

Thank you!