Seneca | SDDS

Business Analytics

Introduction to Data Visualization

BAN140 - Section NBB /NCC

Mufleh Al-Shatnawi, Ph.D., P.Eng.,

Week 4

Week Topics

Previous Week

➢Data Aggregation
➢Basic Data Analysis in Tableau

Current Week

➢Tableau different Join
➢Type of Visuals

Seneca

Introduction

➢ All enterprises/organizations use **data** to monitor operations and perform analysis.

➢ **Analysis are then used to maintain efficiency,**

➢ **Pursue opportunity, and**

➢ **Prevent negative outcomes.**

➢ Rendering data accurately with appropriate visual analytics reduces the time required to achieve understanding.
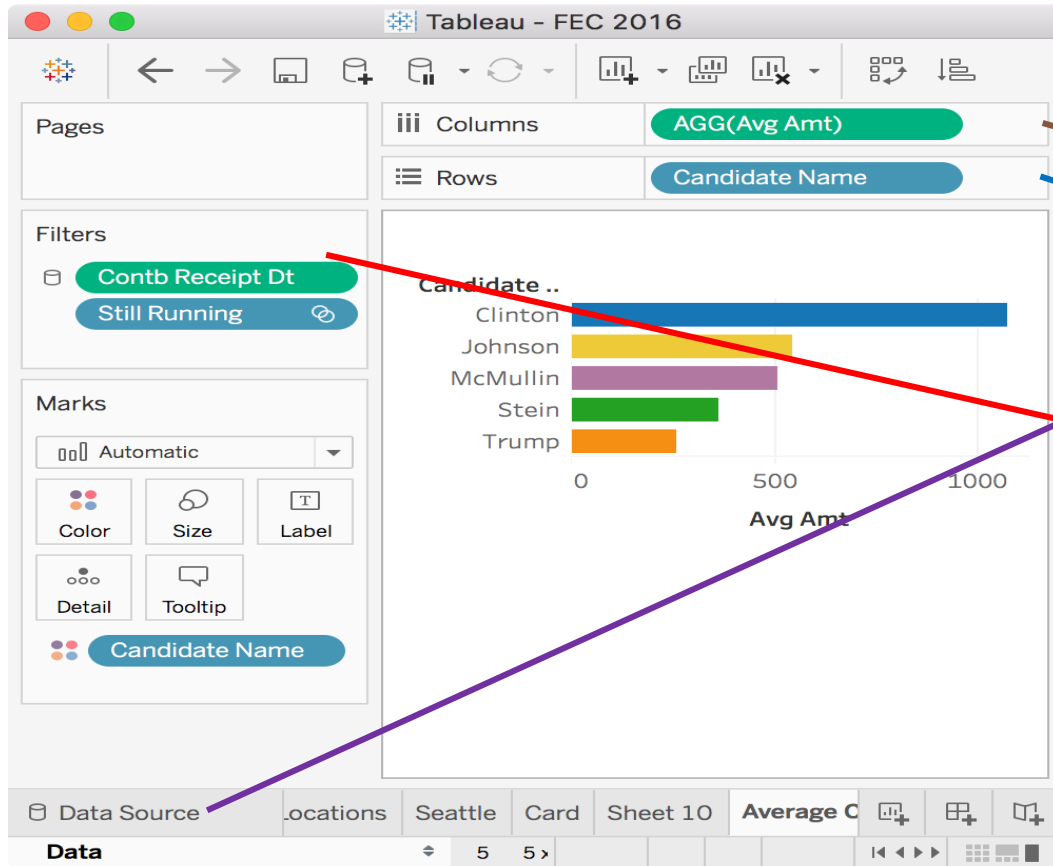
Tableau : Visual Query Language

Tableau is an interface for converting **visual specifications** into **queries**:

SELECT Candidate, AVG(Amount)
FROM FEC
WHERE Date > #2015-01-01#
  AND StillRunning is true
GROUP BY Candidate

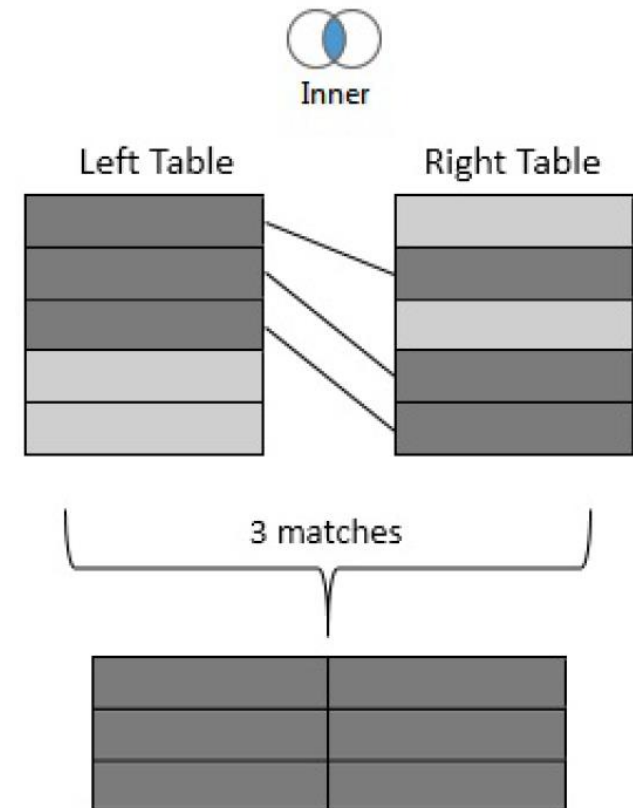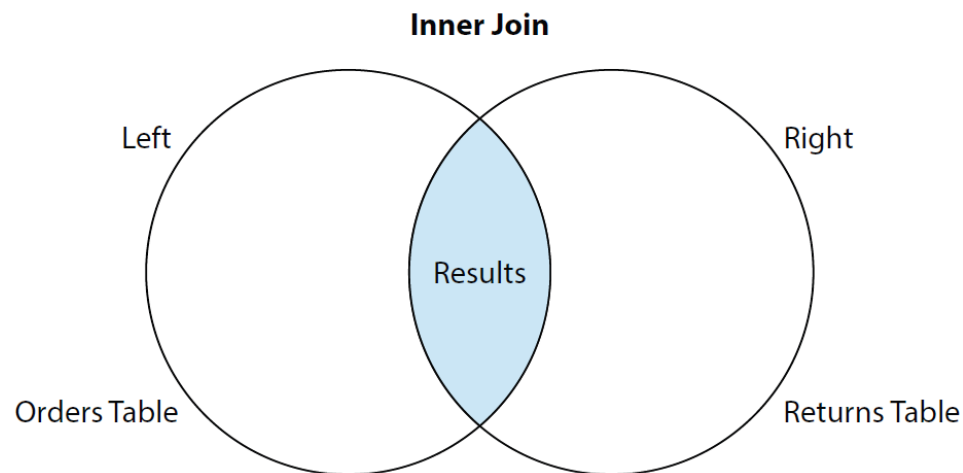Interactivity can generate a **lot of complex queries**

Seneca

Joining Database Tables with Tableau

➢Sometime your datasource does not include every bit of information you need in a single table.

➢So, we can't generate reports using a single table

➢If the data resides in a single spreadsheet or database and each table includes unique identifiers, then we can tie the tables or tabs together.
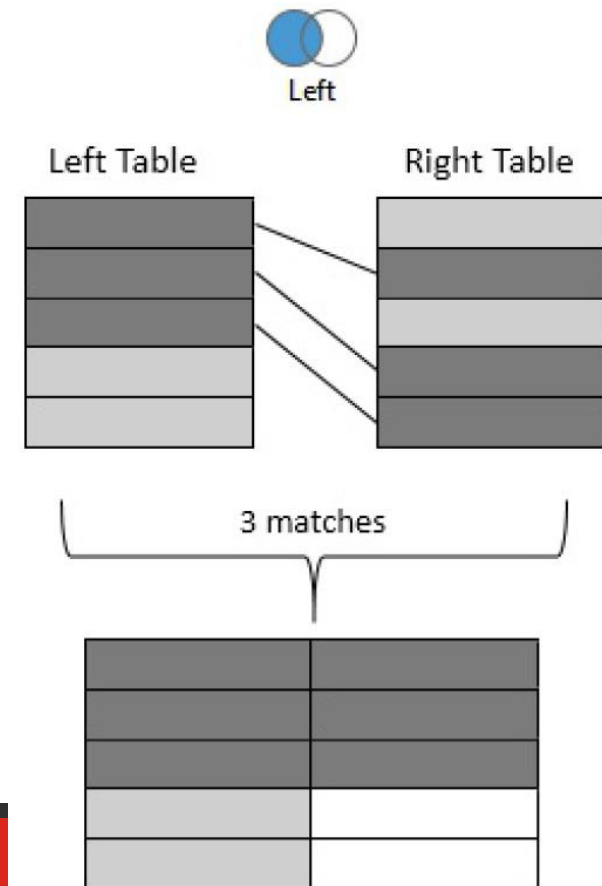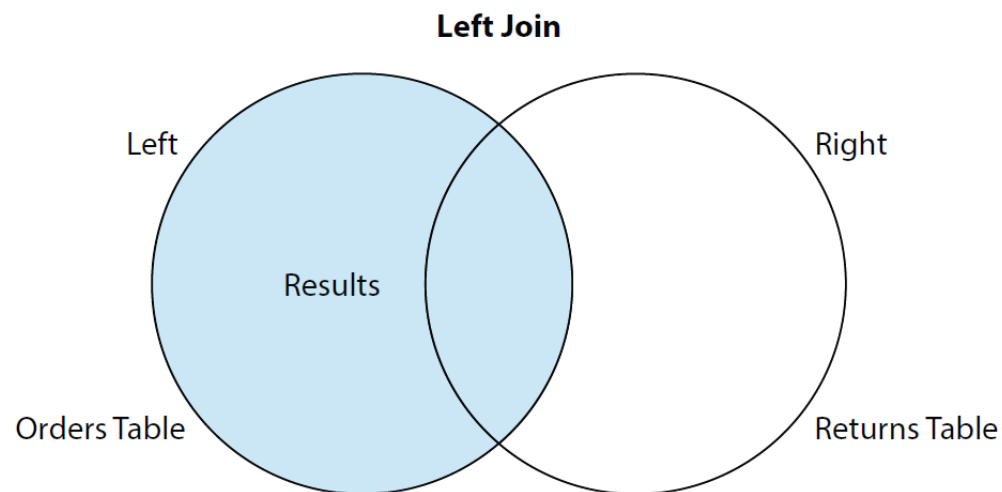
The Default Inner Join

➢**Inner**: Only records that match the join condition from both the table on the left and the table on the right will be kept.
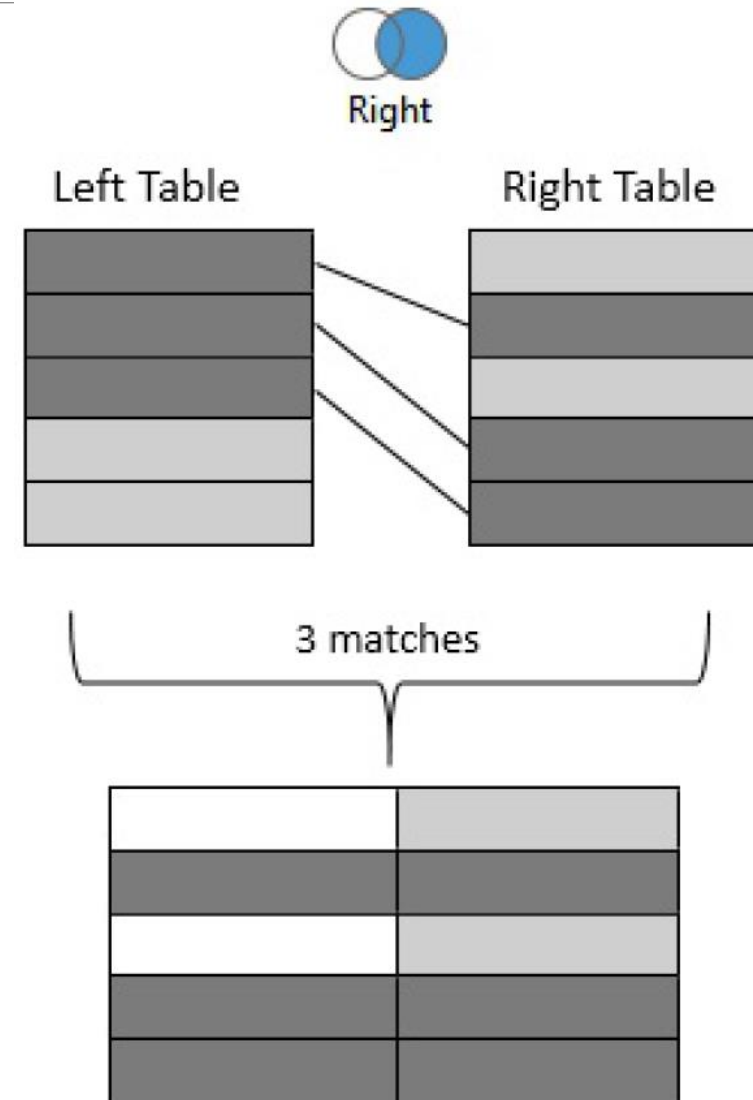
The Left Join Type

➢**Left**: All records from the table on the left will be kept. Matching records from the table on the right will have values in the resulting table, while unmatched records will contain **NULL** values for all fields from the table on the right.

The Right Join Type

➢**Right**: All records from the table on the right will kept. Matching records from the table on the left will result in values, while unmatched records will contain NULL values for all fields from the table on the left. **Not every data source supports a right join.**

Right

Left Table | Right Table

3 matches

The Full Outer Join Type

Full Outer

> **Full Outer**: All records from tables on both sides will be kept. Matching records will have values from the left and the right. Unmatched records will have NULL values where either the left or the right matching record was not found. **Not every data source supports a full outer join**.
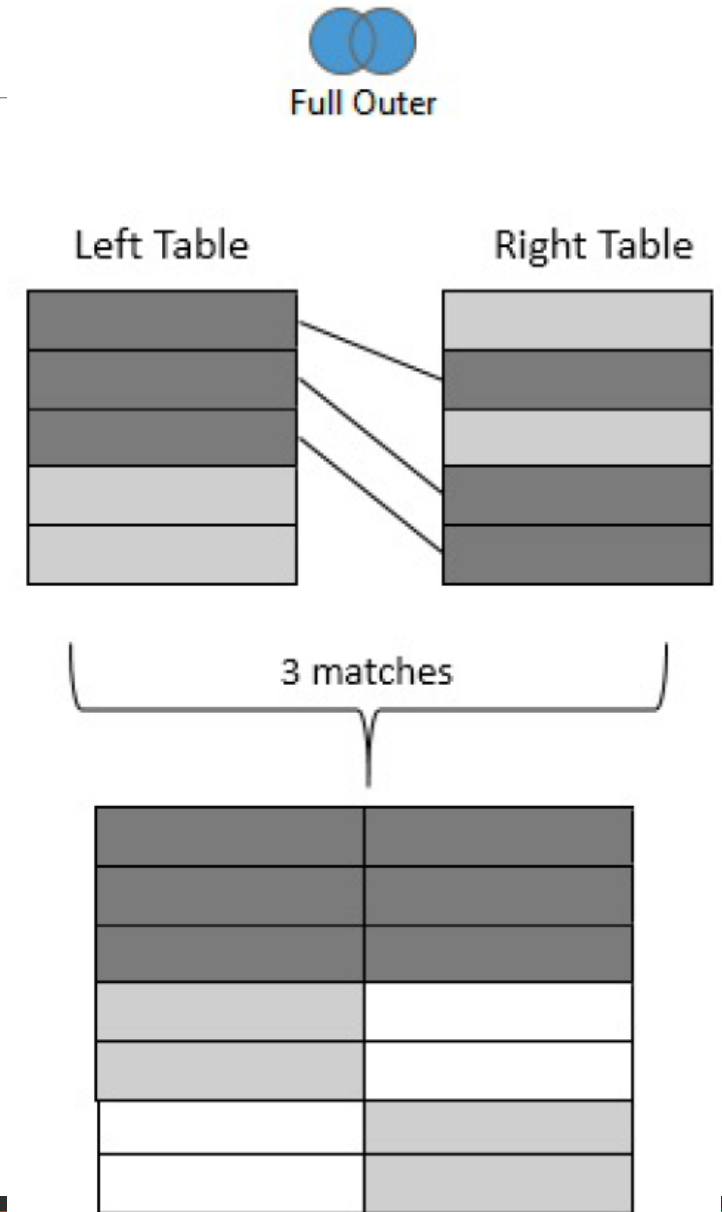
Tableau supports Different join types

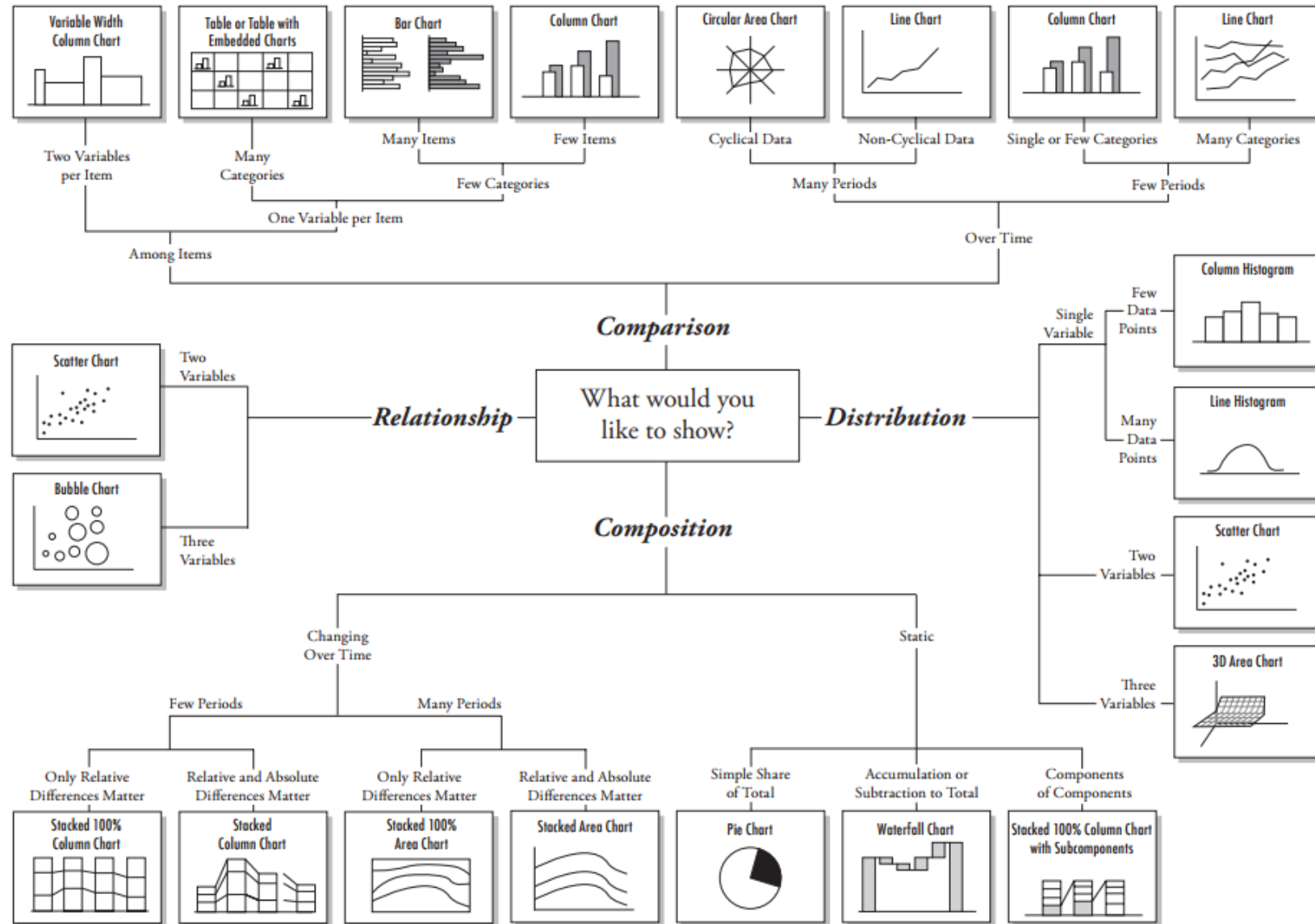| Join Type	Description
Left	For each row, includes all values from the left table and corresponding matches from the right table. When a value in the left table doesn't have a corresponding match in the right table, you see a null value in the join results.
Inner	For each row, includes values that have matches in both tables.
Right	For each row, includes all values from the right table and corresponding matches from the left table. When a value in the right table doesn't have a corresponding match in the left table, you see a null value in the join results.
leftOnly	For each row, includes only values from the left table that don't match any values from the right table. Field values from the right table show as null values in the join results.
rightOnly	For each row, includes only values from the right table that don't match any values from the left table. Field values from the left table show as null values in the join results.
notInner	For each row, includes all of the values from the right and the left table that don't match.
Full	For each row, includes all values from both tables. When a value from either table doesn't have a match with the other table, you see a null value in the join results.

Seneca

Type of Visuals

CHAPTER TWO: CHOOSING AN EFFECTIVE VISUAL FROM STORYTELLING WITH DATA: A DATA VISUALIZATION GUIDE FOR BUSINESS PROFESSIONALS, WILEY, 2017

AVAILABLE ONLINE THROUGH SENECA LIBRARIES: HTTPS://SENECACOLLEGE-PRIMO.HOSTED.EXLIBRISGROUP.COM/PERMALINK/F/T3376V/01SENC_ALMA5146 374280003226

Four questions

1. What data is **important to show**?

2. What do I want to **emphasize** in the data?

3. What **options** do I have for displaying this data?

4. Which option is **most effective** in communicating the data?

Chart Suggestions—A Thought-Starter

© 2006 A. Abela — a.v.abela@gmail.com

Simple Visuals

➢ Many different graphs and visuals are available

➢ However, there are 12 basic ones that are used more often

- Simple Text
- Table
- Heat map

- Scatter plot
- Line
- Slope graph

- Vertical bar
- Horizontal bar

- Stacked vertical bar
- Stacked horizontal bar

- Waterfall
- Square area
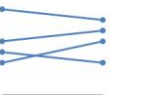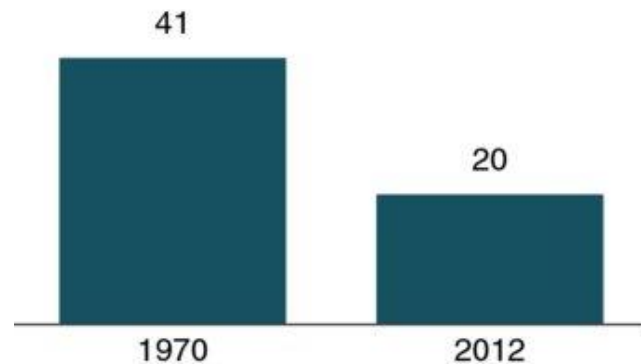
Simple Text (or visual text)

➢When you have just a number or two

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

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

41
20

1970 2012

➡

20%

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

Tables

➤ If you need to communicate multiple different units of measure

➤ Usually not a good idea in a live presentation

➤ Use light or minimal borders, so data stands out (lower ink)

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

Heat map

➢A table with added color saturation

➢Mixes detail with visual cues

➢Always include a legend

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%

From tables to graphs

➢Tables interact with our verbal system

➢Graphs interact with our **visual** system (<u>faster</u>)

➢Graphs
  ➢Points
  ➢Lines
  ➢Bars
  ➢Areas

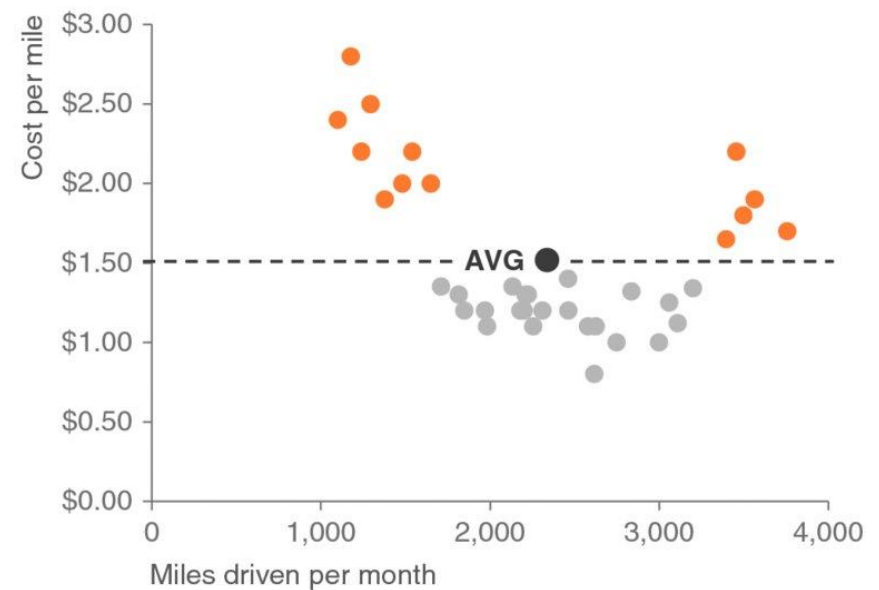# Scatterplot (Points graph)

➢Useful for showing the relationship between two variables

➢Mostly used in scientific fields

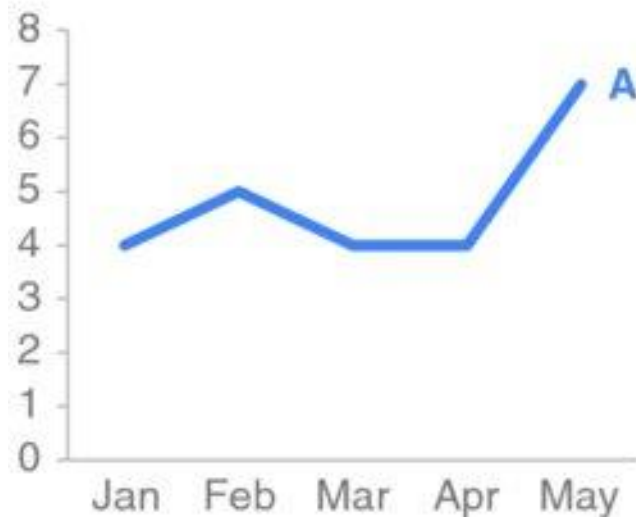➢Use visual cues to draw attention to your story

Line graph

➤ A single <u>series</u> of data (or two or more series)

➤ Some continuation, or trend; e.g. x axis is <u>time</u>

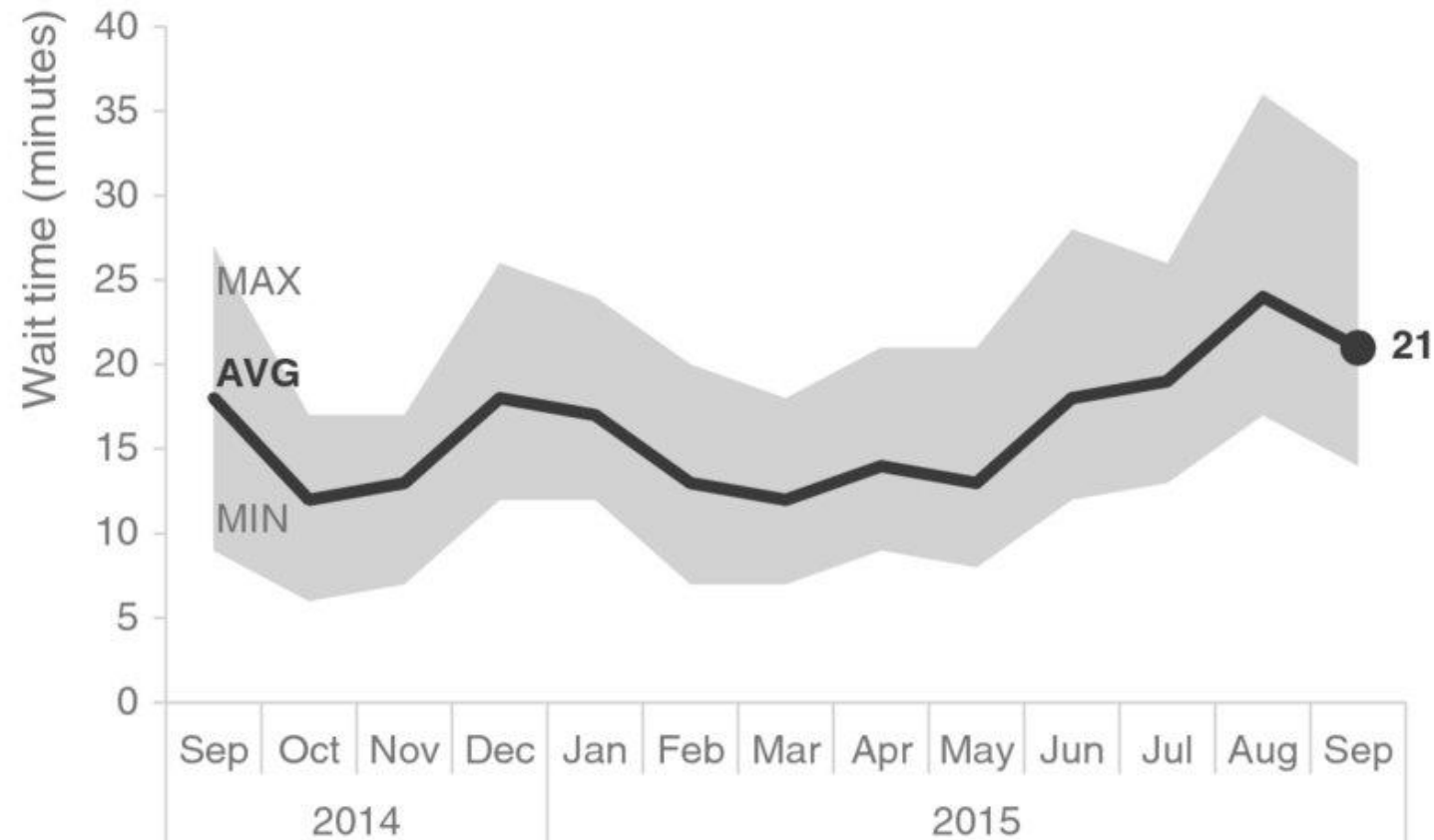➤ Be consistent in the intervals (don't mislead!)

Line graph: Showing average within a range

Passport control wait time
Past 13 months

Slopegraph

➢Two time periods or points of comparison

➢Various categories

➢Show absolute values, as well as the **rate of change (slope)**
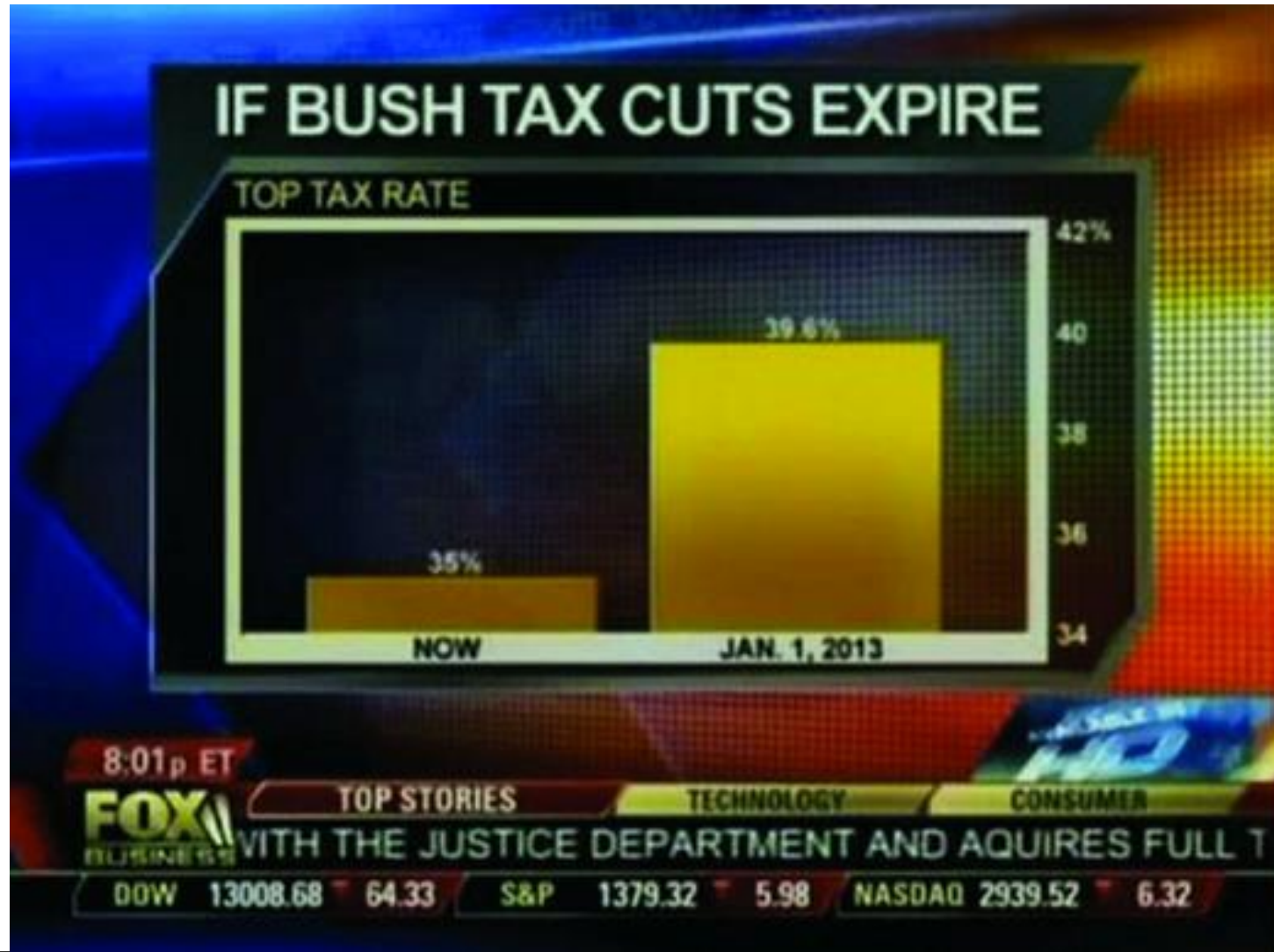
➢Use visual cues to draw attention to your story

Employee feedback over time

Survey category	Percent favorable
Peers	85% → 96%
Culture	80% → 91%
Work environment	76% → 75%
Leadership	59% → 62%
Career development	49% → 33%
Rewards & recognition	41% → 45%
Perf management	33% → 42%

2014   2015

Survey year

Seneca

Bar Charts

➢Suitable when information can be organized into groups

➢Popular and effective

➢Types
  ➢Vertical
  ➢Horizontal
  ➢Stacked vertical
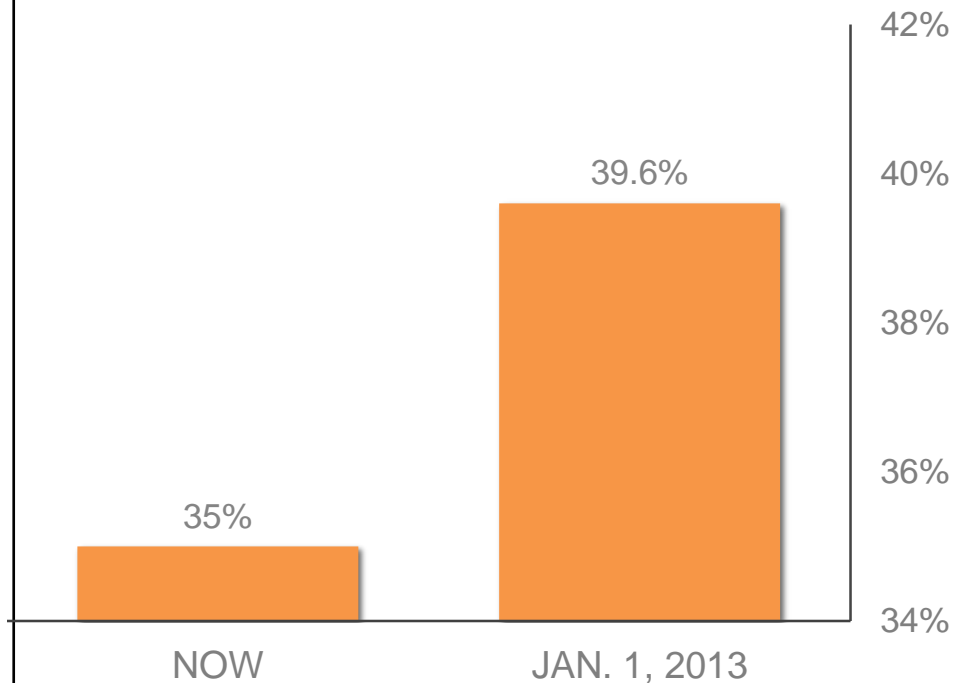 ➢Stacked horizontal
  ➢Waterfall

Common Abuse of Bar Charts
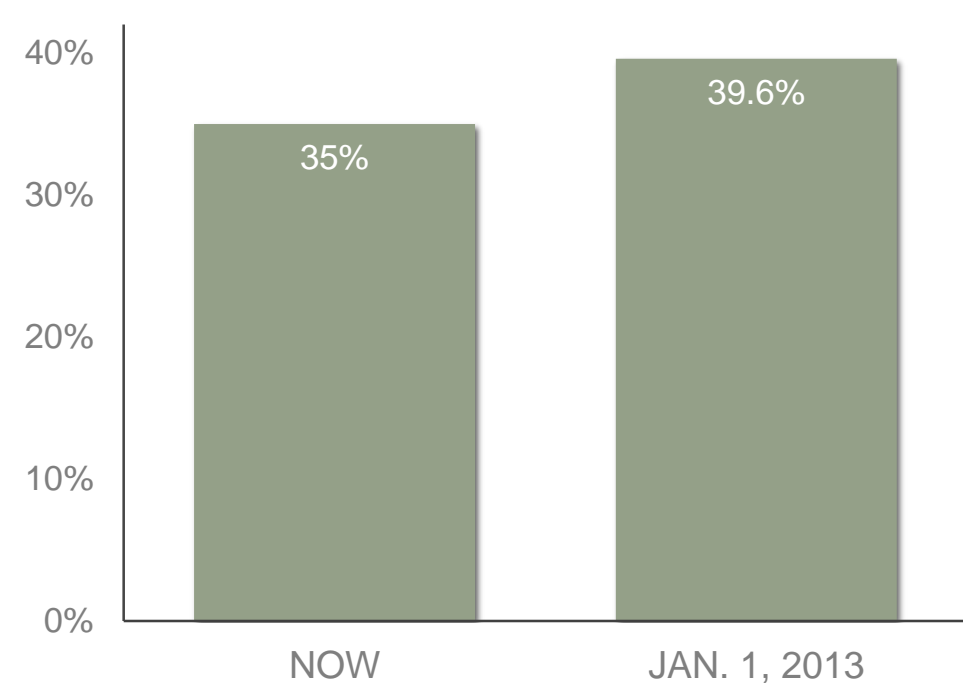
Bar charts must have a zero baseline!

FIG0213

Non-zero baseline: as originally graphed
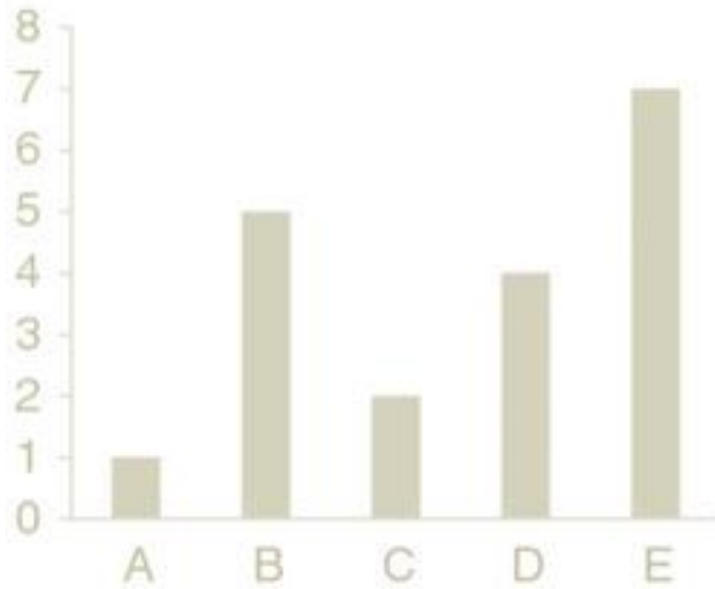
IF BUSH TAX CUTS EXPIRE

TOP TAX RATE

39.6%

35%

NOW   JAN. 1, 2013

42%
40%
38%
36%
34%

Zero baseline: as it should be graphed

IF BUSH TAX CUTS EXPIRE

TOP TAX RATE

39.6%

35%

NOW   JAN. 1, 2013

40%
30%
20%
10%
0%

Width of bars

➢Bars should be wider than the white space between the bars
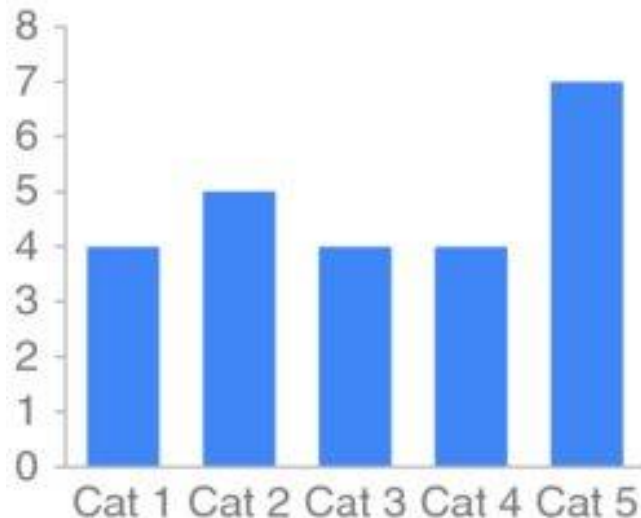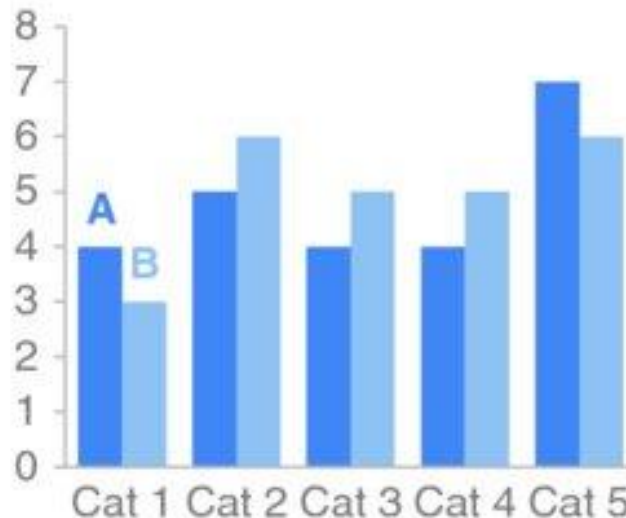
Vertical bar chart

➢Single, two, or multiple series
➢Note it gets harder to focus when many series (avoid!)

➢Visual grouping as a result of spacing
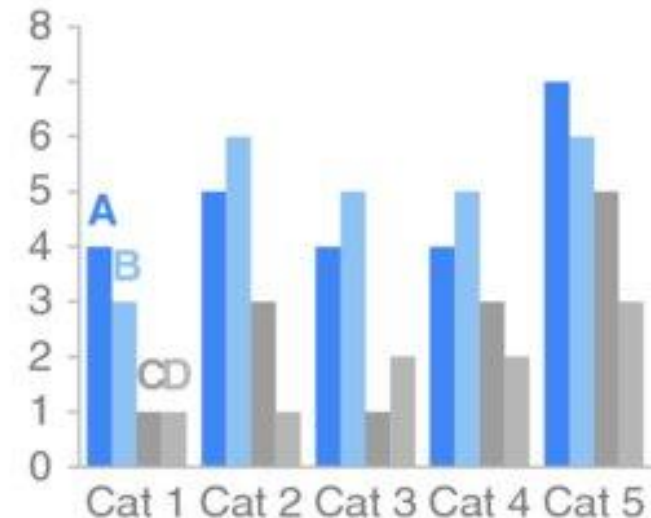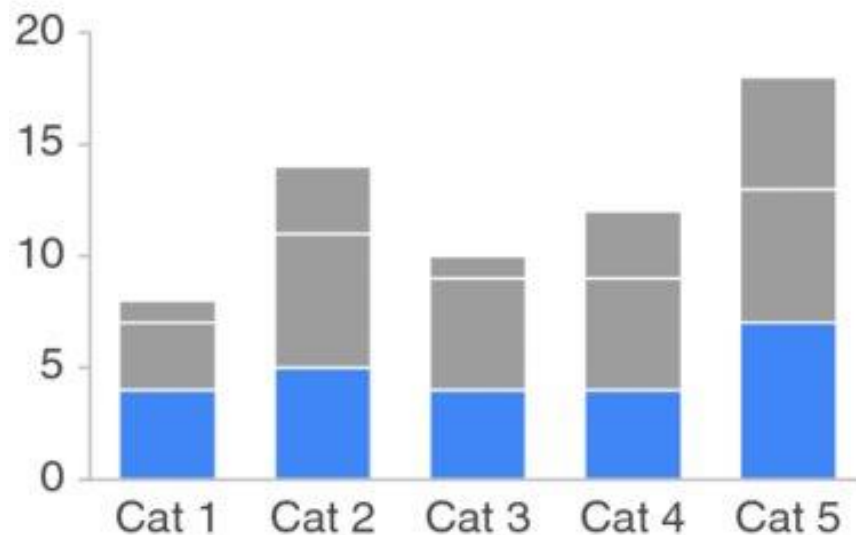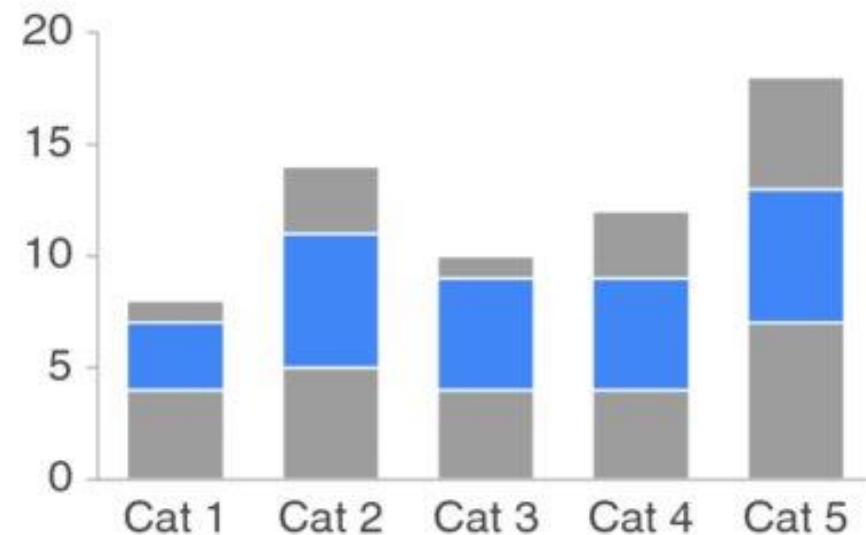➢Pay attention to the relative order (depending on your story)

Stacked vertical bar chart

➢Allows for comparing totals, while seeing subcomponents

➢Can be visually overwhelming

➢Hard to compare subcomponents beyond the bottom one

Comparing **these** is easy
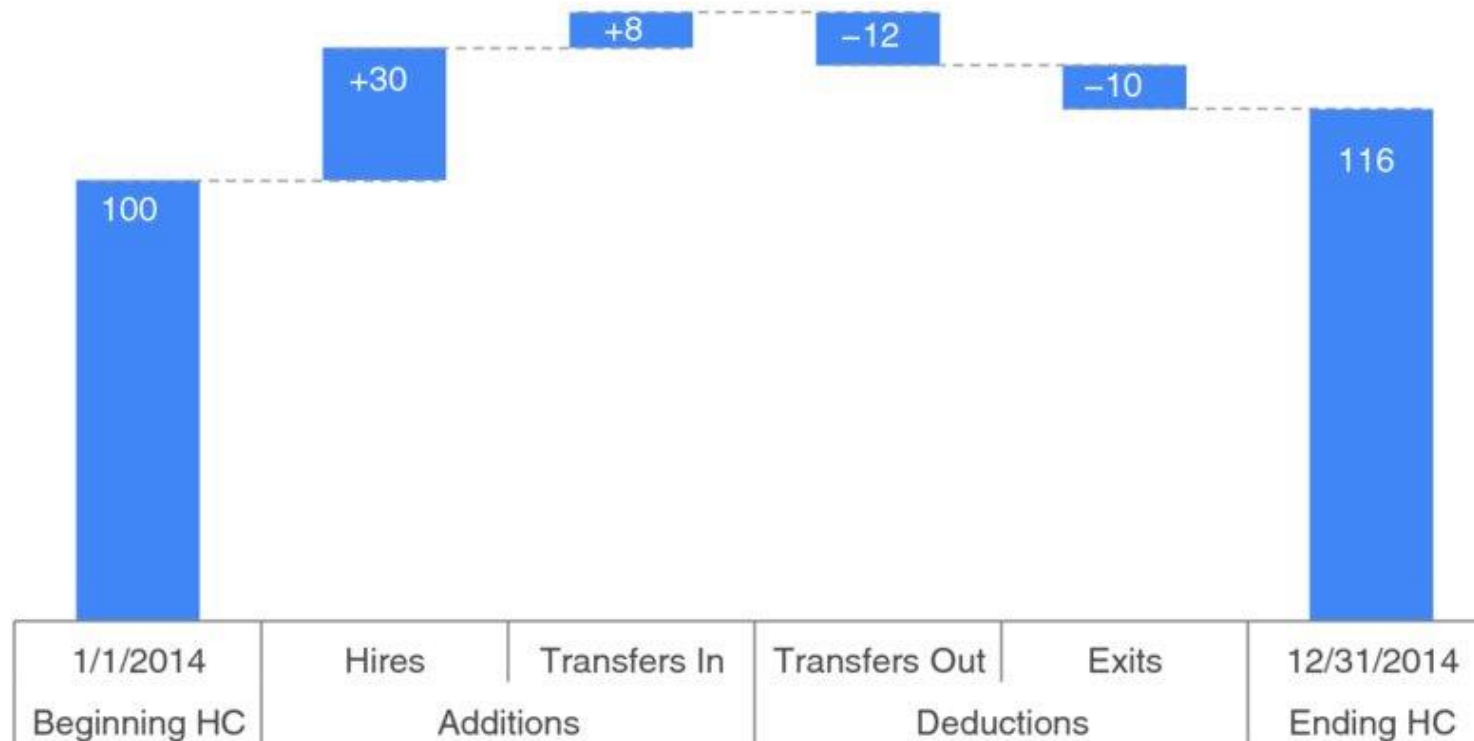
Comparing **these** is hard

Waterfall chart

➢ Shows *changes* (increase or decrease) from a *starting* point to *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.
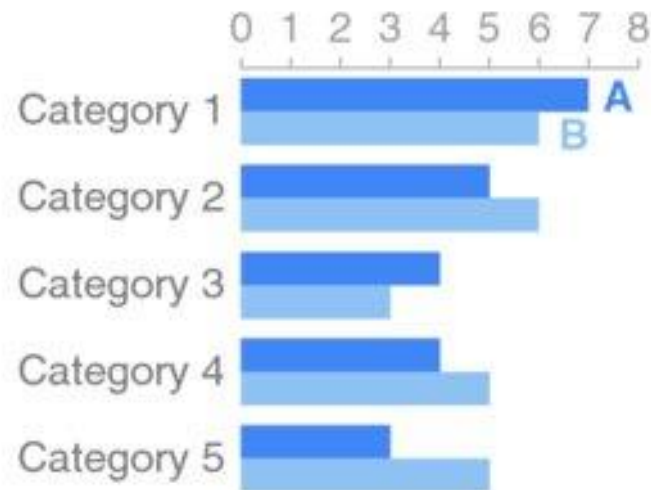
Horizontal bar chart

➤Best option for categorical data

➤Easy to read (left to right, top to bottom, z's)

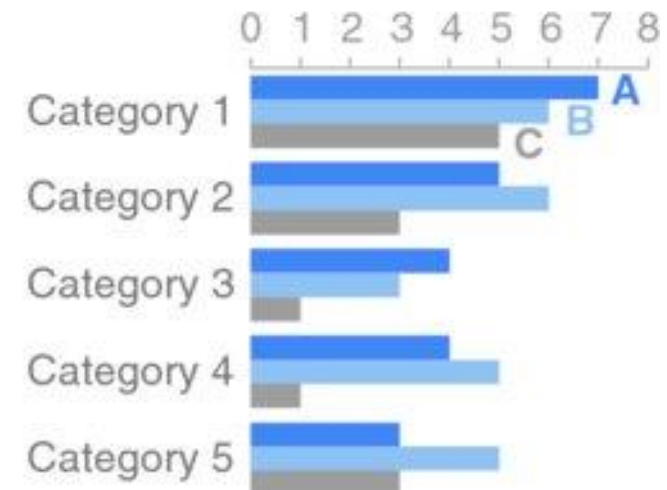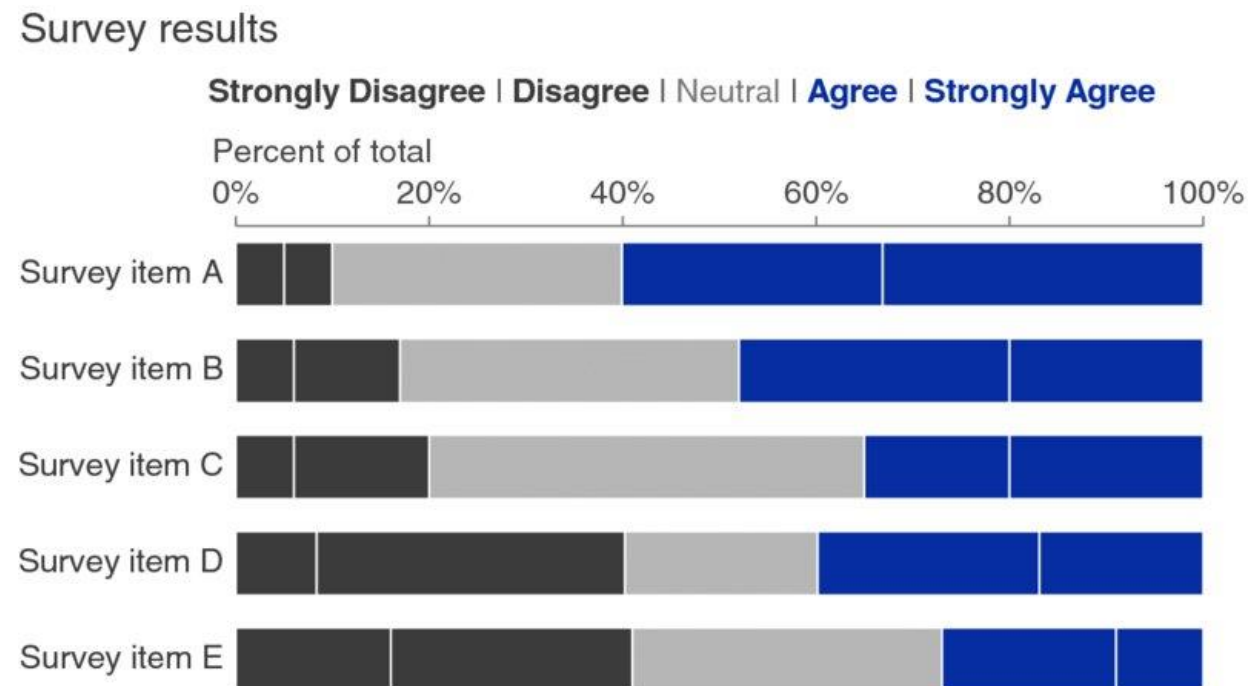➤Note the order of categories; place the important one on top
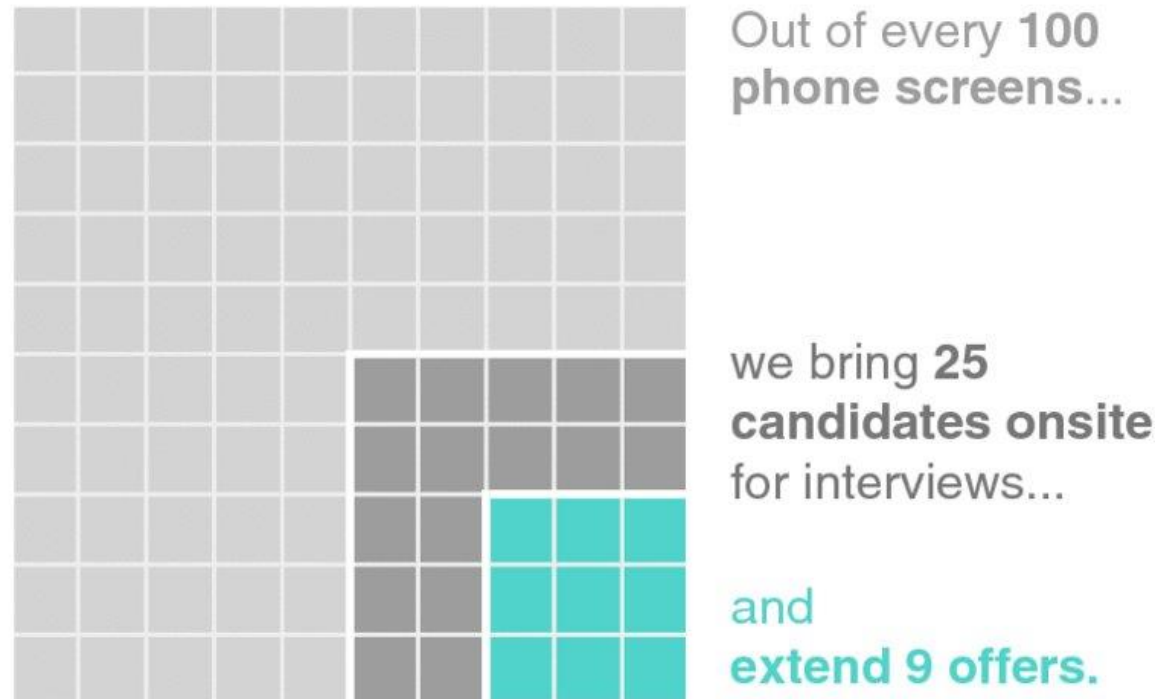
Stacked horizontal bar chart

➢Suitable for either absolute value or when summing up to 100%

➢Used often for visualizing survey data
  ➢*Lickert* scale (strongly agree, agree, etc.)

Survey results

Strongly Disagree | **Disagree** | Neutral | **Agree** | **Strongly Agree**

Percent of total

Area chart

➢When visualizing numbers of vastly different magnitudes
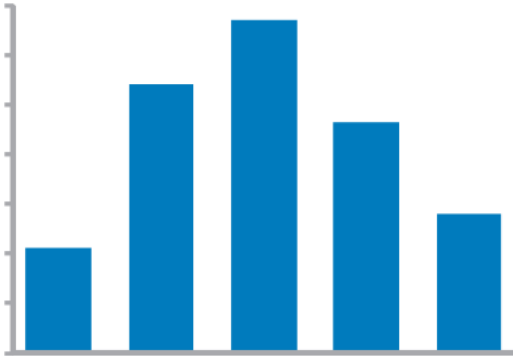
➢Not a recommended method (hard to compare)

Interview breakdown

Out of every **100** phone screens...

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

and extend **9 offers.**

Other Look at how to choose a graph type

Time series
values display how something changed over time

Strip Plot (multiple)
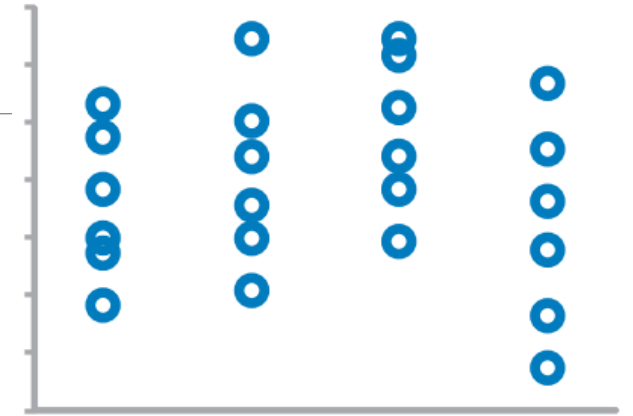
Only when also featuring distributions

Bar Graph (vertical)
To feature individual values and support their comparisons. Quantitative scale must begin at zero.

Line Graph
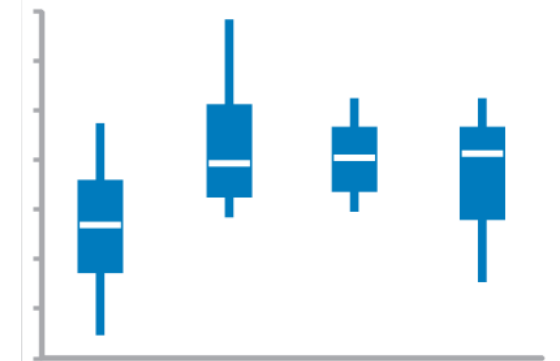To feature overall trends and patterns and support their comparisons

Dot Plot (vertical)
When you do not have a value for every interval of time
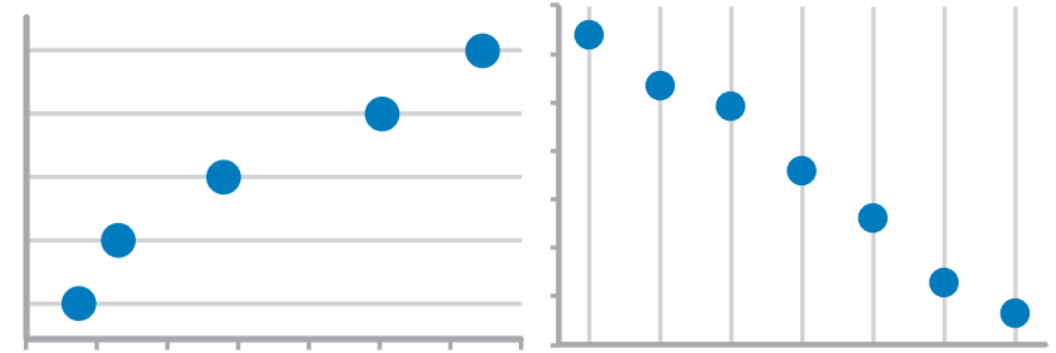
Box Plot (vertical)

Ranking
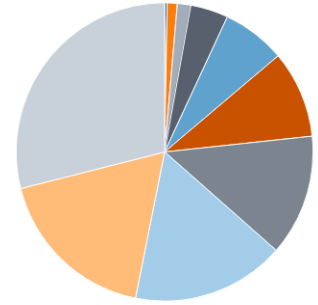## values are ordered by size (descending or ascending)

Bar Graphs

Quantitative scale must begin at zero

Dot Plots

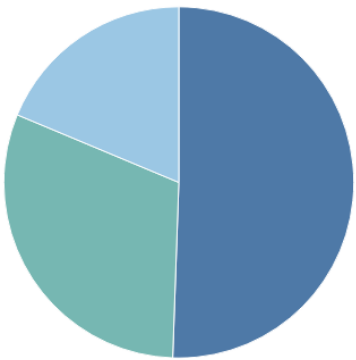# What about pie charts?

➢Commonly used to show parts of a whole

➢However…

 ➢ Hard to judge relative size of pie slices – **better at differentiating length**
 ➢ Take up a lot of space to **present little information**
 ➢**Require labels and good color contrast** to even be usable (often difficult)

Best use is when one overwhelmingly larger value than the rest – no need to focus on actual values

References and Resources

➢[Knaflic] Cole Nussbaumer Knaflic, **Storytelling with Data: A Data Visualization Guide for Business Professionals**, Wiley, 2017
  ➢ Available online through Seneca Libraries: https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA5146374280003226

➢ [Ryan] Lindy Ryan, **Visual Data Storytelling with Tableau**, Pearson Addison-Wesley, 2018
  ➢ Available online through Seneca Libraries: https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA5167006190003226

➢ [Healy] Kieran Healy, **Data Visualization: A Practical Introduction**, Princeton University Press, 2018.
  ➢ Available (hardcopy) at Seneca Libraries: https://senecacollege-primo.hosted.exlibrisgroup.com/permalink/f/t3376v/01SENC_ALMA2172469250003226

➢**A Reader on Data Visualization:** https://mschermann.github.io/data_viz_reader/

➢**Data visualization:** https://en.wikipedia.org/wiki/Data_visualization

➢**Section 5: Data concepts** https://www.statcan.gc.ca/eng/dli/guide/section5