DATA SOCIETY®

The premiere data science training for professionals

Welcome!

- Instructor introduction
- Schedule
  - Mon-Thurs: 9 – 2 pm for instructional hours
  - Mon-Thurs: 2 – 4 pm for independent work
  - Friday: independent work / capstone work
  - Holidays: September 6 & 9 (Labor Day), and October 14 (Columbus Day)
- Attendance
- Student introductions

Program overview

- By the end of this program, you will be able to:

1. Program in multiple languages such as R, Python, and SQL
2. Build powerful predictive models to identify patterns and anticipate trends
3. Build data dashboards and intuitive visualizations to communicate findings effectively
4. Demonstrate your skills with final capstone projects

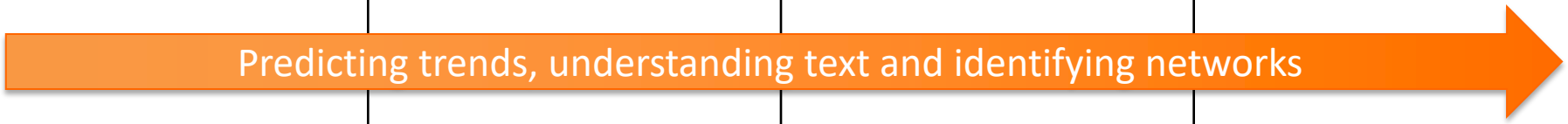
Topics overview

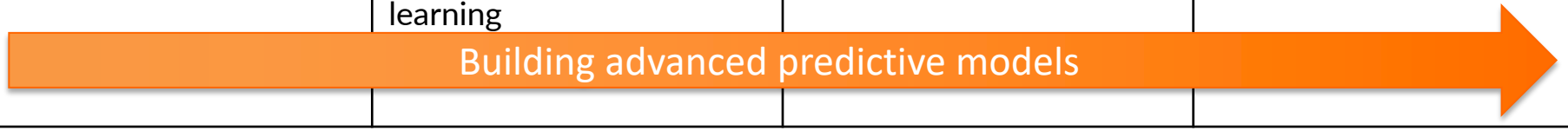
AFWERX program roadmap

INTRO TO DATA SCIENCE	R PROGRAMMING	POWER BI	PYTHON PROGRAMMING	SQL	MACHINE LEARNING
• What is data science? • How can you become data-driven? • What type of problems can data science solve?	• What is R? • How is R used as a tool for data science? • How can you visualize data with R?	• What is Power BI? • How can you build an effective data dashboard? • How can you effectively communicate your results?	• What is Python? • How is Python used as a tool for data science? • How can you visualize data with Python?	• How can you store structured data in a SQL database? • How can you query the database to understand the data?	• What are the most useful supervised and unsupervised techniques? • How can you use machine learning to solve problems?

Program schedule

Introduction to data science	Introduction to R programming	Interactive visualizations and RShiny	Power BI
Introduction to Python	Advanced visualization with Python	Introduction to SQL	Clustering
Regression and time-series analysis	Advanced classification	Text mining & sentiment analysis	Introduction to Network Graph Analysis
Community detection	Introduction to neural networks and deep learning	Supervised deep learning methods	Capstone reviews and final projects





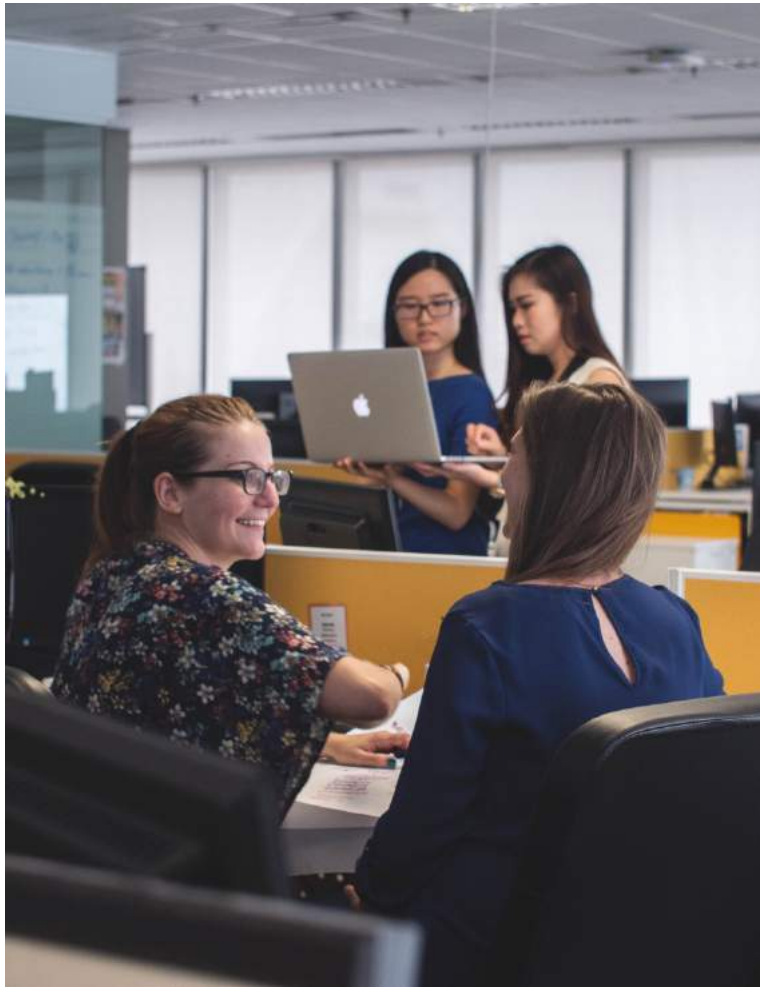
Capstone objectives

1. Build a practical application that solves a use case for your team
2. Mine a data set for patterns
3. Build models that reveal insights or provide forecasts
4. Create visualizations and dashboards that tell the story you want to convey

You'll be presenting your project at the end of the program

Best practices for success!

1. Be ready to start and return from breaks on time
2. Please put phones away
3. Participate in activities and ask questions
4. Dedicate time to capstone projects



Intro activity: are you data-driven?

- Turn to your participant guide to the **Data-driven assessment** page to evaluate your team

- Once you've answered the questions and identified your quadrant, discuss your results with your group – are there any trends that you've noticed? What are some key areas that you are interested in improving?

Activity time: 15 - 20 minutes

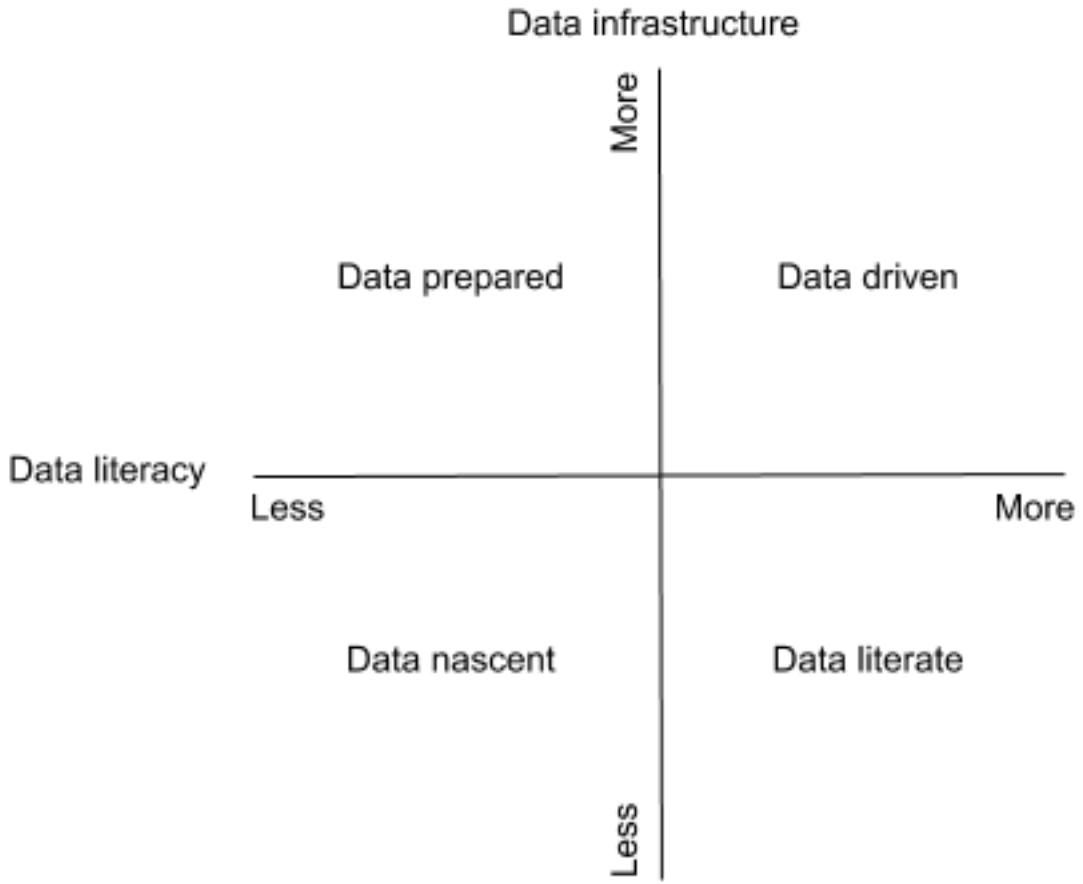
Data-driven metrics

Data infrastructure metrics:

1. Data access
2. Data storage
3. Data collection

Data literacy metrics:

1. Data leadership
2. Data governance
3. Staff knowledge of data



Outline for today

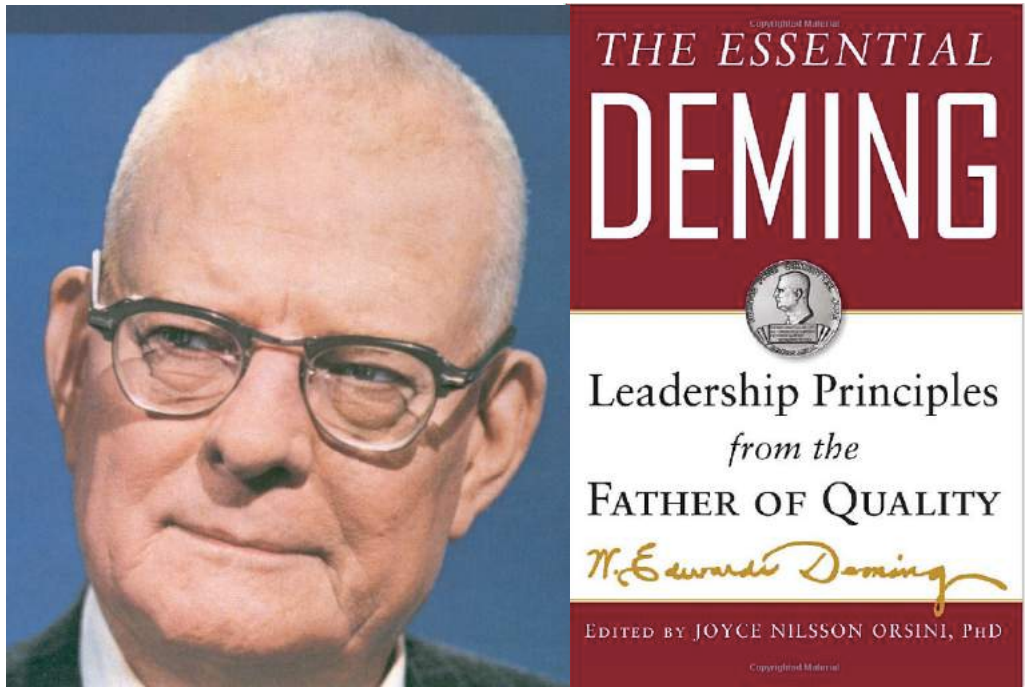
1. Data driven benefits
2. Overview of data science methods
3. Data visualization and storytelling
4. Data ethics frameworks
5. Open data sources

Data vs. a hunch

In God we trust. All others must bring data.

- William Edwards Deming, 1950
(Professor / consultant famous for work in quality control)

- Brought to Japan by General Douglas MacArthur to fix Japan's infrastructure after WWII
- Deming trained hundreds of engineers, managers, and scholars in statistical process control (SPC) and concepts of quality
- He is largely credited with the success of quality control at Sony, Toyota and Honda



Definitions: data vs. no data

- What is worse than being wrong? Not being accountable!
- When experts are wrong, they're rarely held accountable and rarely admit it
- Common reasons for being wrong:
  - Just off on timing
  - Blindsided by an "improbable event(s)"
  - Almost right
  - Wrong for the right reasons
- No more inclined than anyone else to revise their beliefs about the way the world works because they made a mistake

Instrumental value of data

- Data may be retained for a number of reasons:
  - Compliance: avoiding penalties
  - Automation: economic efficiencies
  - Low Value Analytics: naive insights, and sometimes paradoxical ones
  - High Value Analytics:** insights that are nontrivial, novel, and comprehensive

6 key benefits of data driven strategies

1. Find a needle in haystack

2. Prioritize work for high impact

3. Early warning / detection

4. Faster decisions

5. Optimize resources

6. Experiment for what works

Strategy 1: find a needle in haystack

- Governments can now use these strategies and predictive analytics to better target their activities

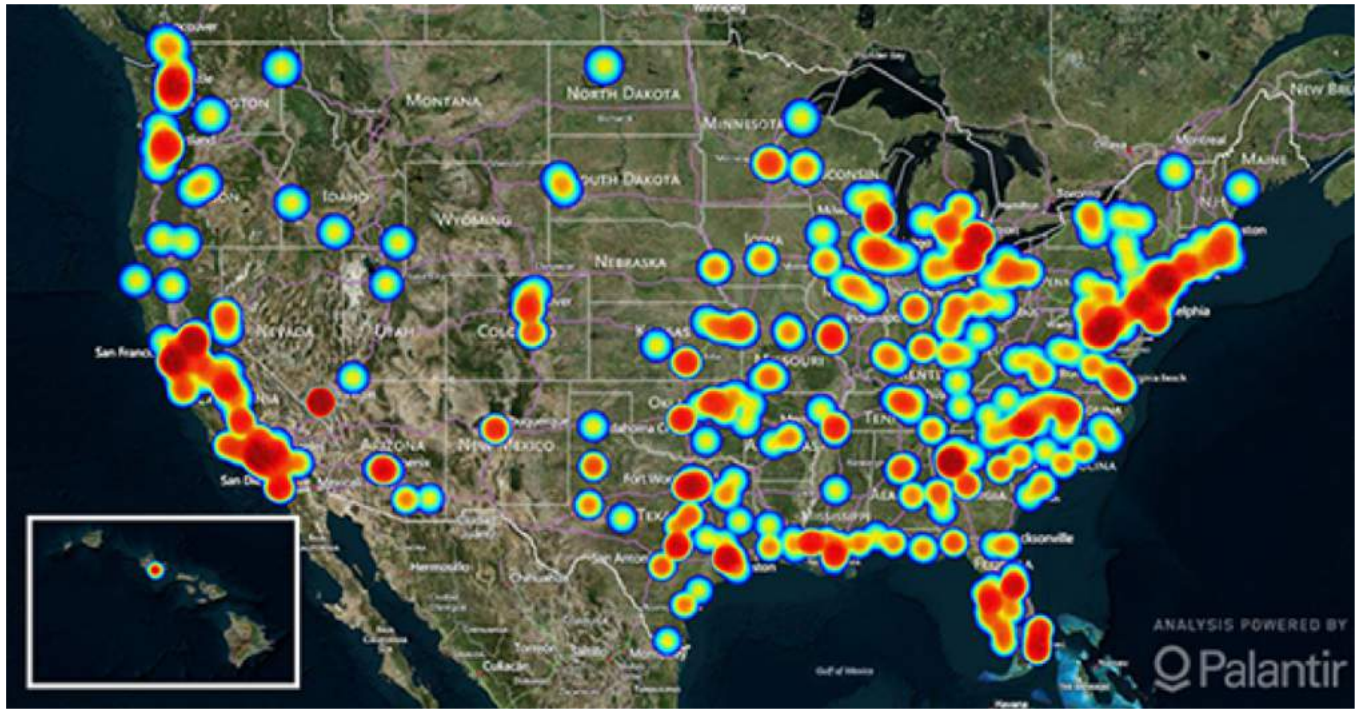
Business application	Public sector application
A process involves a large number of similar decisions, but each decision requires considering a wide range of variables.	Inspections
The problem and the outcome involve a system and not just one agency (i.e., where treating the manifestation of the problem may miss much of the underlying cause).	Truancy (public safety and education)
There is abundant information in electronic data form available on which to base decisions and measure outcomes	Fraud detection
It is possible to insert a model calculation into the actual business process, either to automate decisions or to support human decision makers.	Welfare casework
Conditions of privacy, standardization, data security, and transparency have been explicitly considered.	Tax collection

Strategy 1: find a needle in haystack

- Challenge: targets are difficult to identify or locate within a broader population
- Opportunity: data analysis and predictive modeling to identify targets based on existing data

Human trafficking predictive analytics

- Polaris in Washington DC is targeting and visualizing massage parlors because they have made a connection to human trafficking



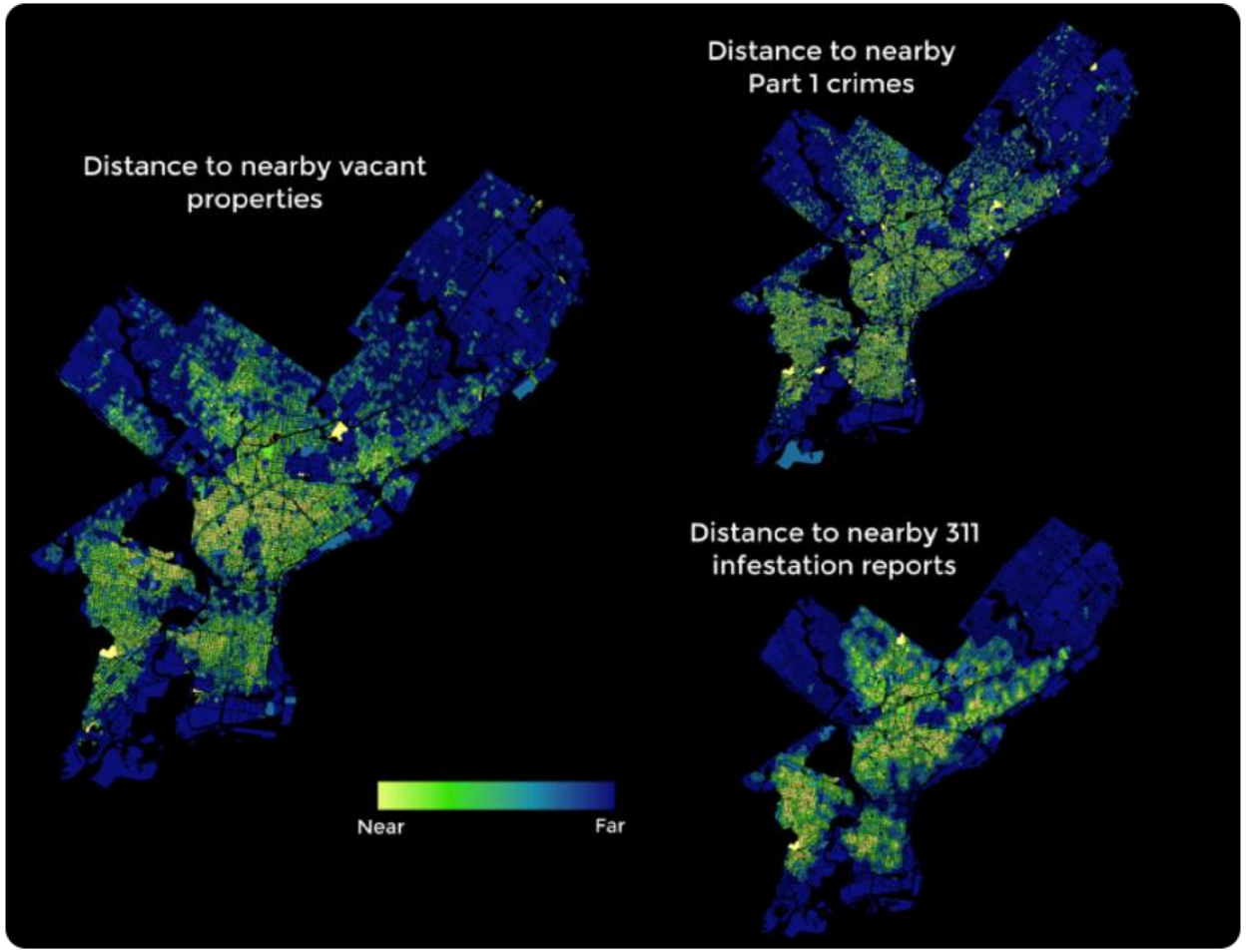
https://www.datanami.com/2016/10/07/data-analytics-fight-human-trafficking/

Strategy 2: prioritize work for high impact

- Challenge: services do not categorize high-priority cases early
- Opportunity: data analysis and predictive modeling to prioritize cases

Building safety

- Philadelphia - predictive modeling to prioritize building inspections
- Models estimate the probability that a given building might be unsafe



http://urbanspatialanalysis.com/portfolio/proof-of-concept-using-predictive-modeling-to-prioritize-building-inspections/

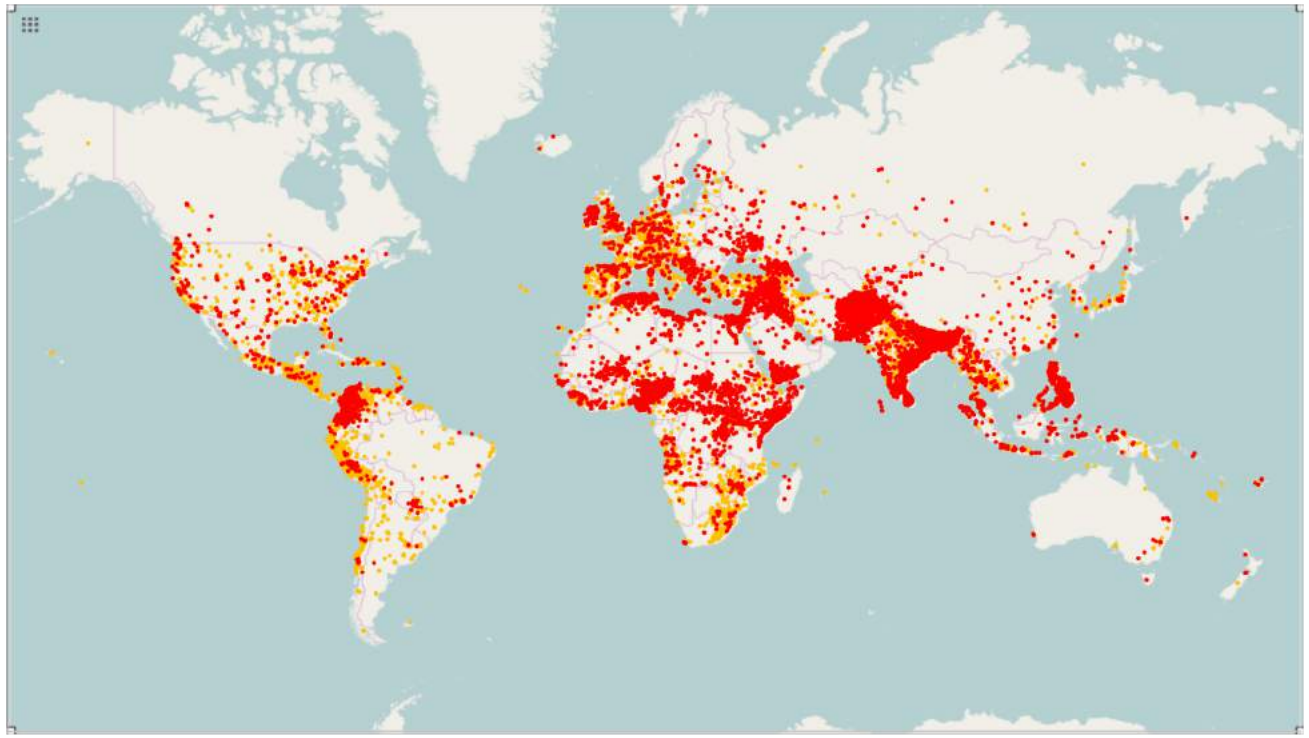
Strategy 3: early warning detection

- Challenge: resources overly focused on reactive services
- Opportunity: develop tools to predict need based on historic patterns

Terrorism - early warning tools

- Why do terrorist attacks frequently succeed, even though later investigations almost always show that warnings had been available but were either misunderstood or ignored?

- Government agencies are now more careful in utilizing technologies for planning and preparing for attacks. A new framework developed by researchers at Binghamton University, State University of New York is able to understand future terrorist behaviors by recognizing patterns in past attacks with 90% accuracy



https://www.sciencedaily.com/releases/2017/03/170302115740.htm

Strategy 4: better, quicker decisions

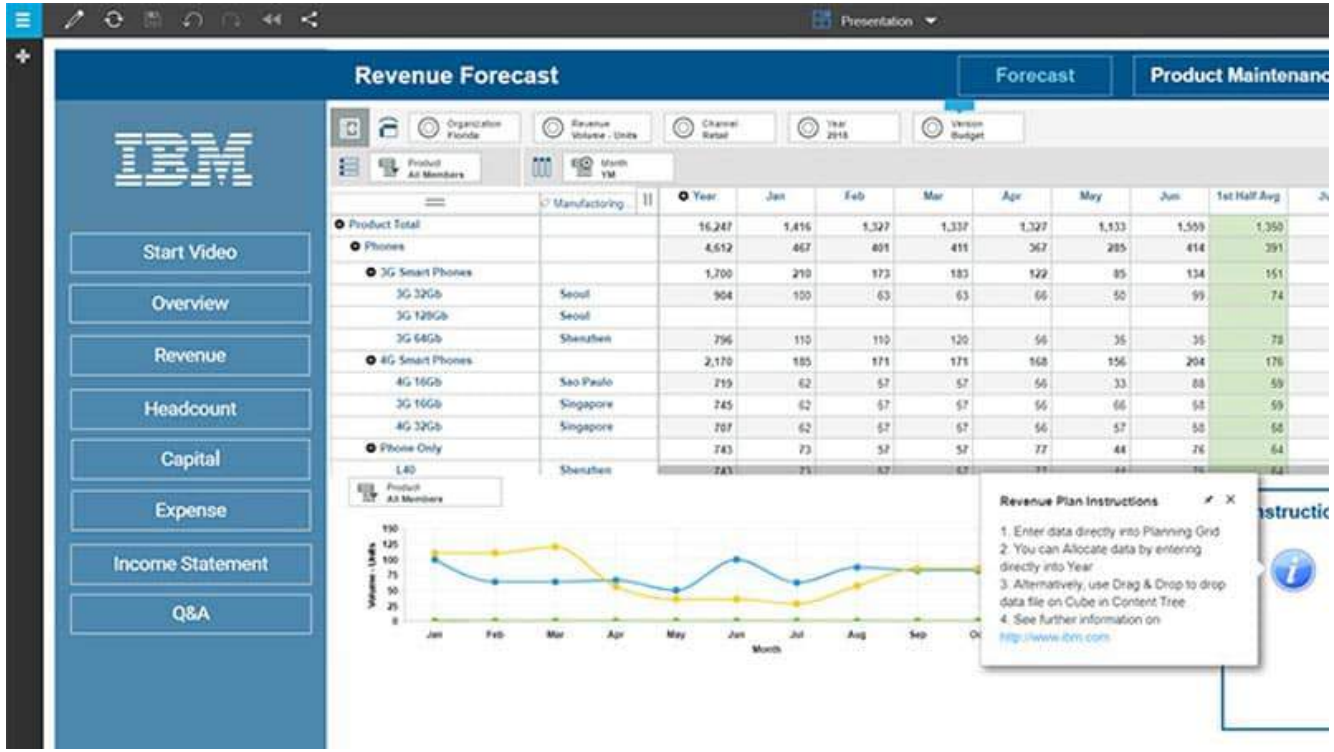
- Challenge: repeated decisions are made without access to all relevant information
- Opportunity: develop recommendation tools for operational decisions

Government – better, quicker decisions

- A large federal intelligence agency had more than a dozen homegrown financial databases. Analysts spent many hours entering data and manipulating spreadsheets to create a consolidated view of the government agency's spending

- Solution- consolidate government agency budget data on one analytics platform (this agency used IBMs planning analytics platform)

- The federal government agency was able to reduce their annual actual execution reconciliation process from three weeks to one day



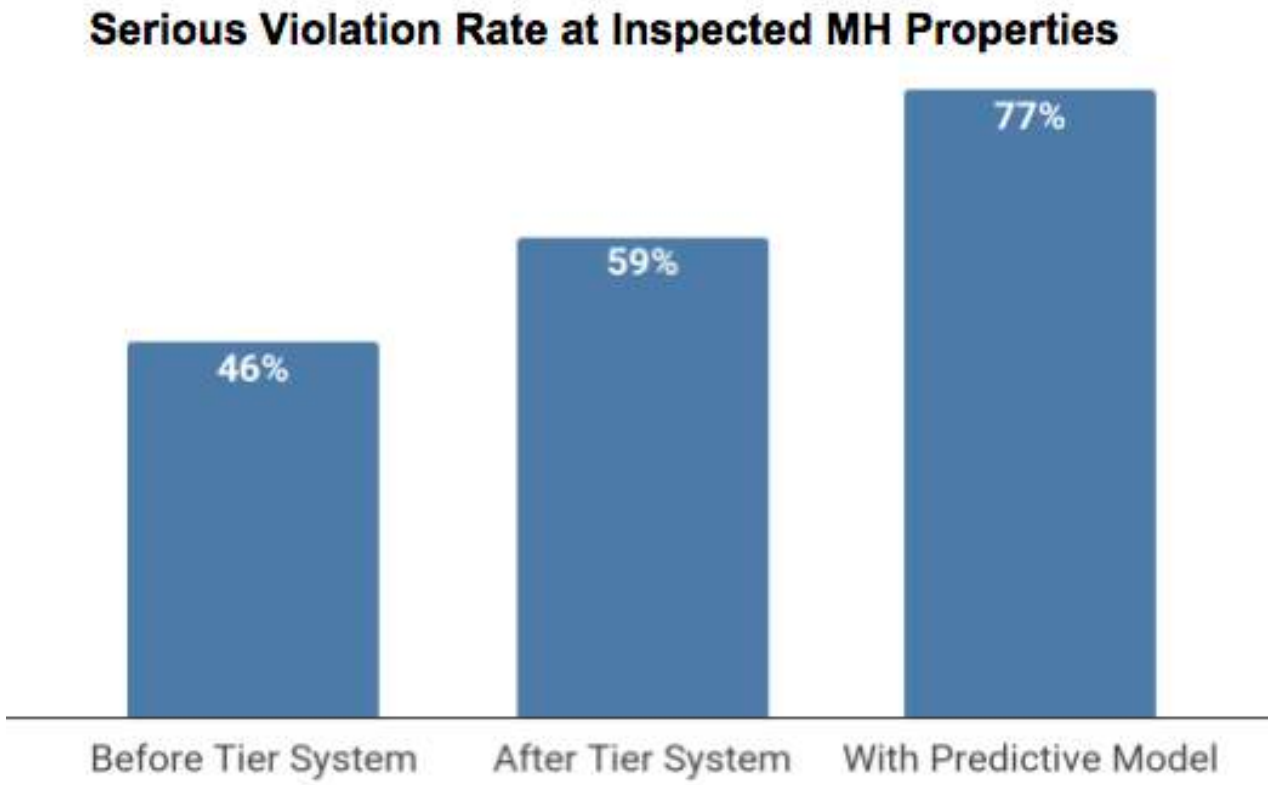
https://www2.deloitte.com/us/en/pages/deloitte-analytics/articles/Government-agency-data-analytics.html

Strategy 5: optimize resources

- Challenge: assets are scheduled or deployed without input of latest service data
- Opportunity: use data to drive decisions on the deployment of resources

Prioritizing building inspections

- In June 2015, a balcony collapsed in Berkeley, killing six people and injuring seven more
- San Jose's Multiple Housing Team implemented data analytics to prioritize high-risk cases
- By identifying the most important factors, they built a model to help prevent future building collapses



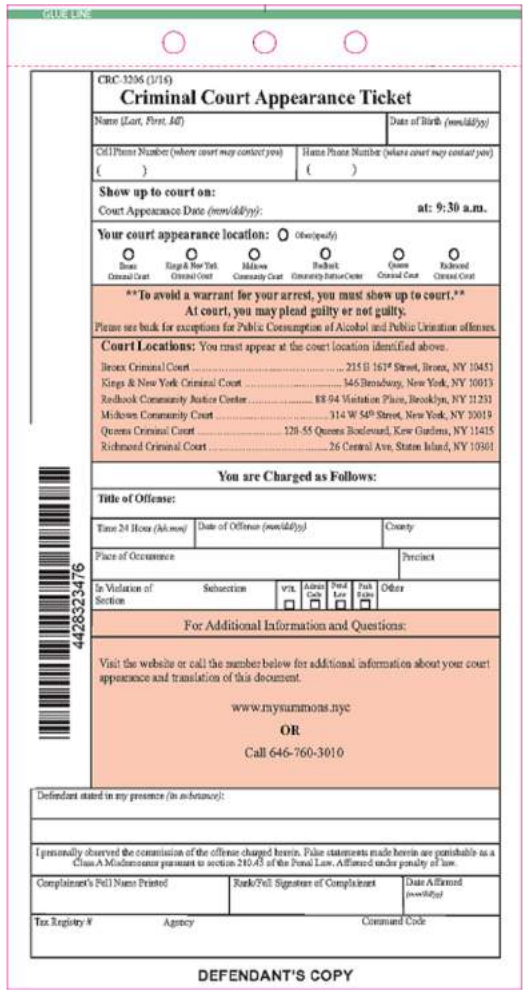
https://dssg.uchicago.edu/2017/07/14/data-driven-inspections-for-safer-housing-in-san-jose-california/

Strategy 6: experiment for what works

- Challenge: services have not been assessed for impact
- Opportunity: experimental testing and improvement of service options

Form redesign

- The NYC government redesigned its court summons notices to reduce the number of people who fail to appear (FTA)
- The redesign alone corresponded to a drop in FTAs by 13%
- When paired with text messaging, there was a 36% decrease!



CRC-3206 (1/16)

Criminal Court Appearance Ticket

Name (*Last, First, MI*) | Date of Birth (*mm/dd/yy*)

Cell Phone Number (*where court may contact you*) | Home Phone Number (*where court may contact you*)

Show up to court on:

Court Appearance Date (*mm/dd/yy*): **at: 9:30 a.m.**

Your court appearance location: Other (*specify*)

Bronx Criminal Court | Kings & New York Criminal Court | Midtown Community Court | Redhook Community Justice Center | Queens Criminal Court | Richmond Criminal Court

****To avoid a warrant for your arrest, you must show up to court.****
At court, you may plead guilty or not guilty.
Please see back for exceptions for Public Consumption of Alcohol and Public Urination offenses.

Court Locations: You must appear at the court location identified above.

Bronx Criminal Court ..... 215 E 161st Street, Bronx, NY 10451
Kings & New York Criminal Court ..... 346 Broadway, New York, NY 10013
Redhook Community Justice Center ..... 88-94 Visitation Place, Brooklyn, NY 11231
Midtown Community Court ..... 314 W 54th Street, New York, NY 10019
Queens Criminal Court ..... 120-55 Queens Boulevard, Kew Gardens, NY 11415
Richmond Criminal Court ..... 26 Central Ave, Staten Island, NY 10301

You are Charged as Follows:

Title of Offense:

Time 24 Hour (*hh:mm*) | Date of Offense (*mm/dd/yy*) | County

Place of Occurrence | Precinct

In Violation of Section | Subsection | VTL | Admin Code | Penal Law | Park Rules | Other

For Additional Information and Questions:

Visit the website or call the number below for additional information about your court appearance and translation of this document.

www.nysummons.nyc
OR
Call 646-760-3010

Defendant stated in my presence (*in substance*):

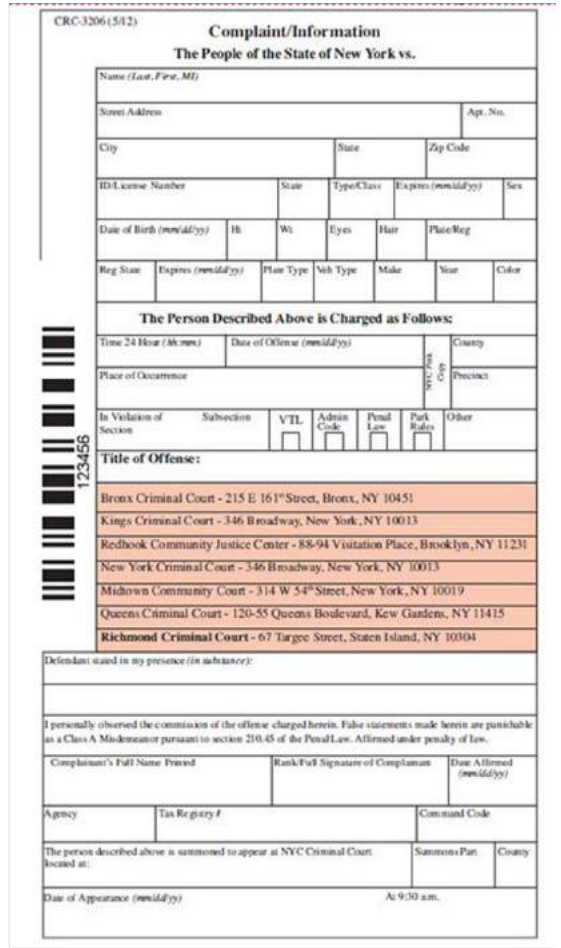
I personally observed the commission of the offense charged herein. False statements made herein are punishable as a Class A Misdemeanor pursuant to section 210.45 of the Penal Law. Affirmed under penalty of law.

Complainant's Full Name Printed | Rank/Full Signature of Complainant | Date Affirmed (*mm/dd/yy*)

Tax Registry # | Agency | Command Code

DEFENDANT'S COPY

442823476



CRC-3206 (5/12)

Complaint/Information

The People of the State of New York vs.

Name (*Last, First, MI*)

Street Address | Apt. No.

City | State | Zip Code

ID/License Number | State | Type/Class | Expires (*mm/dd/yy*) | Sex

Date of Birth (*mm/dd/yy*) | Ht | Wt | Eyes | Hair | Plate/Reg

Reg State | Expires (*mm/dd/yy*) | Plate Type | Veh Type | Make | Year | Color

The Person Described Above is Charged as Follows:

Time 24 Hour (*hh:mm*) | Date of Offense (*mm/dd/yy*) | NYC Code | County

Place of Occurrence | Precinct

In Violation of Section | Subsection | VTL | Admin Code | Penal Law | Park Rules | Other

Title of Offense:

Bronx Criminal Court - 215 E 161st Street, Bronx, NY 10451
Kings Criminal Court - 346 Broadway, New York, NY 10013
Redhook Community Justice Center - 88-94 Visitation Place, Brooklyn, NY 11231
New York Criminal Court - 346 Broadway, New York, NY 10013
Midtown Community Court - 314 W 54th Street, New York, NY 10019
Queens Criminal Court - 120-55 Queens Boulevard, Kew Gardens, NY 11415
Richmond Criminal Court - 67 Targee Street, Staten Island, NY 10304

Defendant stated in my presence (*in substance*):

I personally observed the commission of the offense charged herein. False statements made herein are punishable as a Class A Misdemeanor pursuant to section 210.45 of the Penal Law. Affirmed under penalty of law.

Complainant's Full Name Printed | Rank/Full Signature of Complainant | Date Affirmed (*mm/dd/yy*)

Agency | Tax Registry # | Command Code

The person described above is summoned to appear at NYC Criminal Court located at: | Summons Part | County

Date of Appearance (*mm/dd/yy*) | At 9:30 a.m.

123456

http://www.ideas42.org/wp-content/uploads/2018/03/Using-Behavioral-Science-to-Improve-Criminal-Justice-Outcomes.pdf

What made these projects successful?

1. A clear objective with specific scope: this can be set by the project manager in conjunction with team members

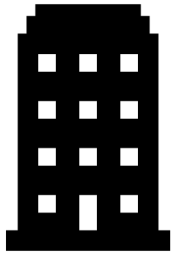
2. Manager-level support and resources: managers should ensure that there is sufficient budget, and that team members have the tools and backing to pursue the project

3. Knowledgeable team members: your team should understand the problem, as well as how to implement the correct algorithms to identify solutions

4. A consistent flow of accurate data: without data, none of this would be possible!

What types of data could you use?

- What data do you currently have access to?



Internal

Publicly available

Social media

One of the most powerful aspects of data analysis is in combining disparate data sources to find new patterns and insights!

Activity: brainstorm ideas!

- Turn to your participant guide to the Project brainstorm page to identify 3- 5 ideas that would improve processes on your team or in the Air Force
- Write each on a separate Post It note (the questions and ideas should be related to the strategies discussed)
- Then, discuss your ideas with the class, and write them down in the appropriate quadrant based on feasibility and impact

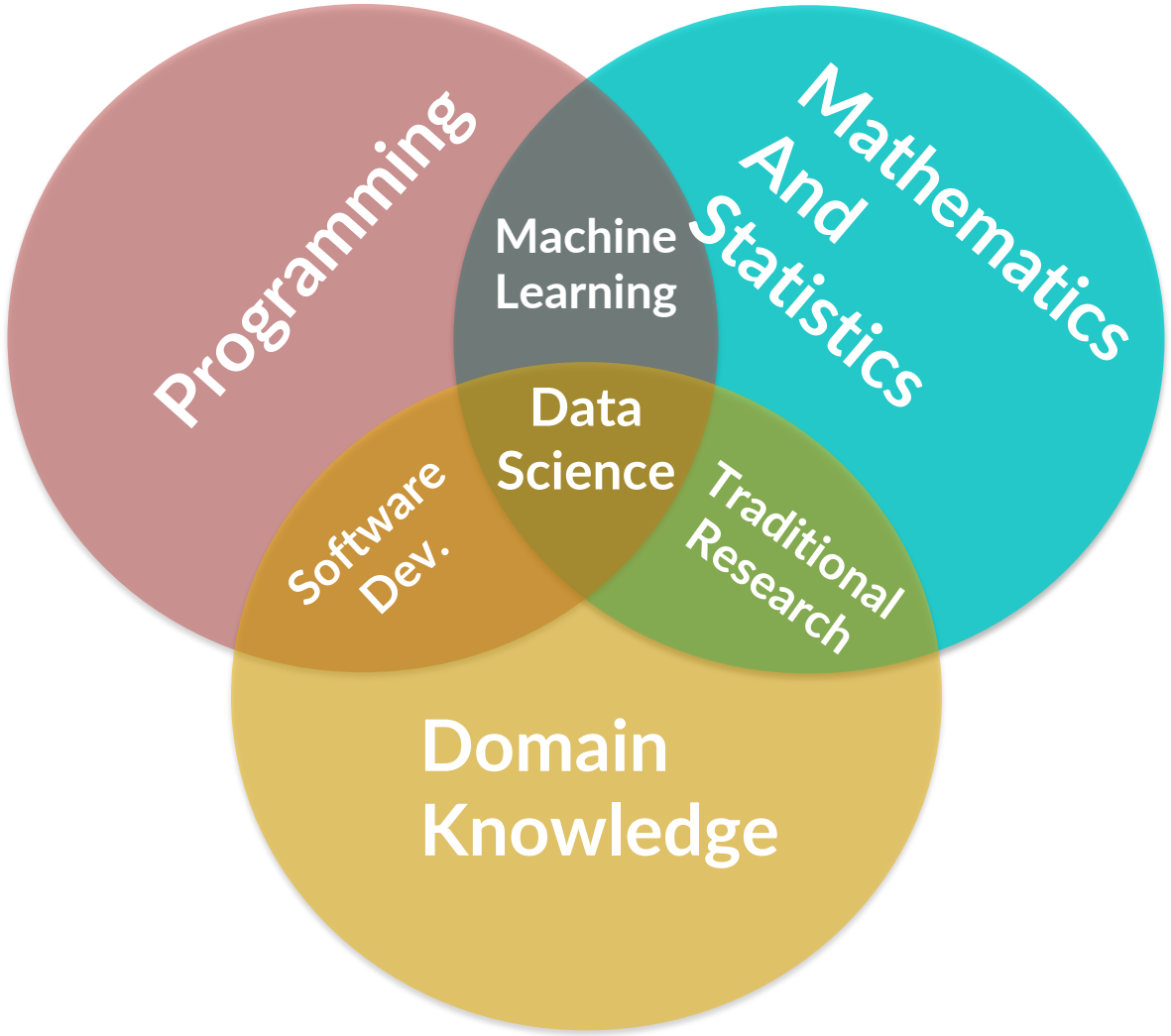
Activity time: 15 - 20 minutes

Outline for today

1. Data driven benefits
2. Overview of data science methods
3. Data visualization and storytelling
4. Data ethics frameworks
5. Open data sources

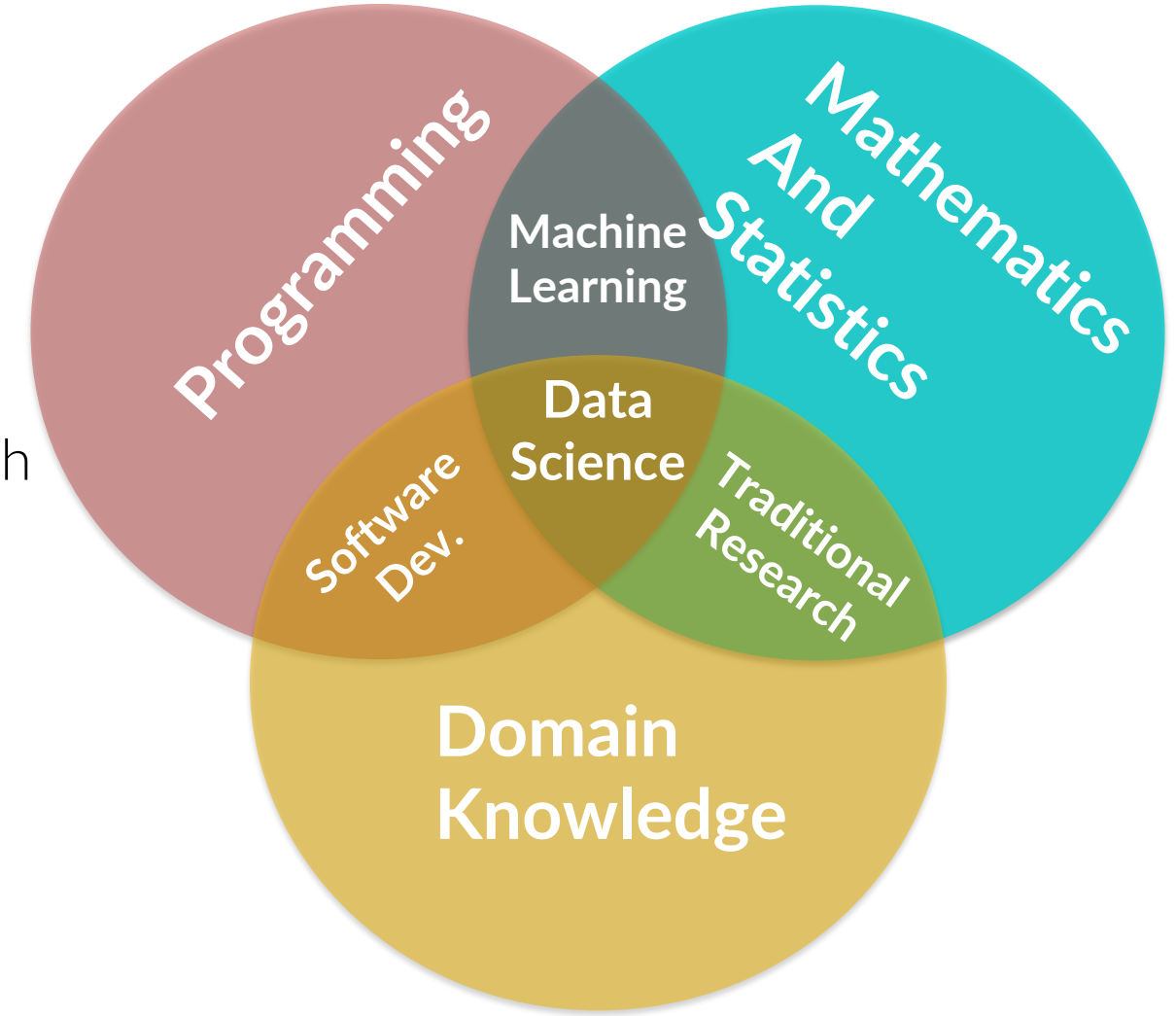
What is data science?

- Data science applies the scientific method to analyzing data
- It lies at the intersection of several disciplines
- It draws on domain specific knowledge that makes the analysis of Big Data possible



Data science for managers

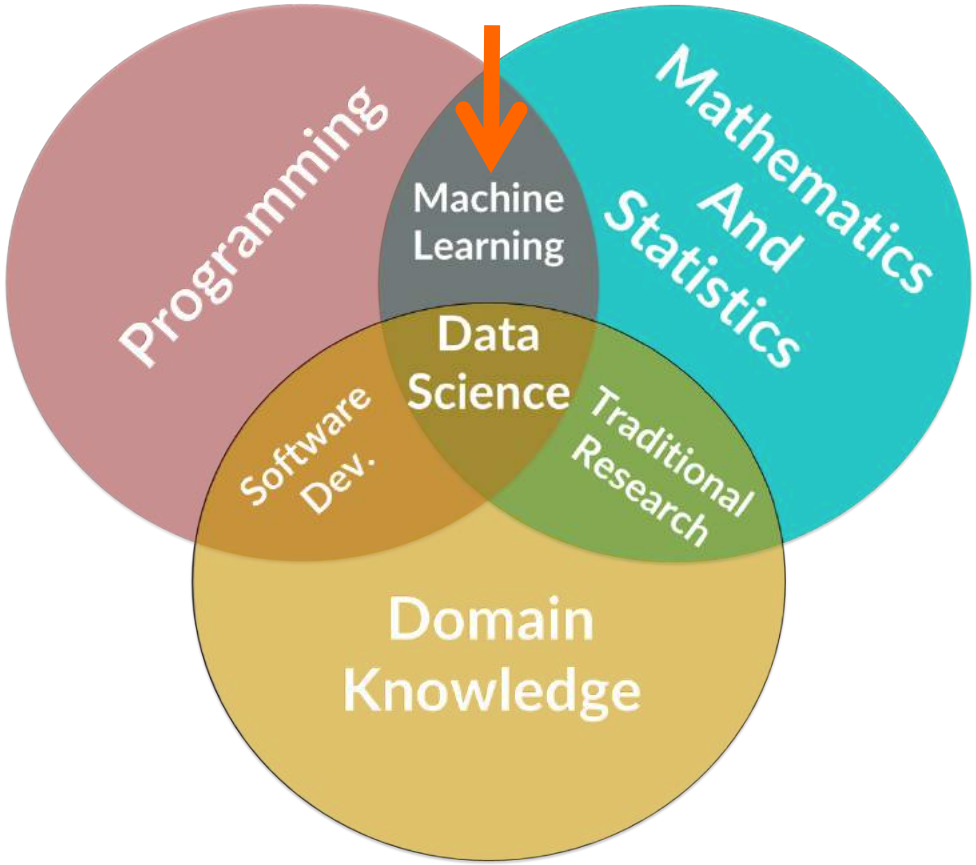
- Data science is principles, not algorithms!
- Domain knowledge is essential to knowing what to look for when working with data
- Correct interpretation is the difference between good data science and using English treasure discoveries to predict stock prices
- Managers have to understand the fundamental principles well enough to envision and appreciate data science opportunities



Setting the record straight on buzzwords

- Machine learning
- Artificial intelligence
- Statistical learning
- Neural Networks
- Deep Learning

Computers recognizing patterns to help make more accurate forecasts, and help make decisions with better results



What is "Big Data"?

- Big data refers to a large volume of structured, semi-structured, and unstructured data that can be mined for information and used in machine learning projects and other analytics applications

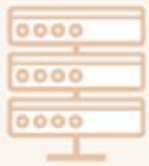
- Typically, the size of big data is described in terabytes, petabytes, even exabytes!

- "Big data" *is not* analytics – you can't "do" big data, but you can use it for data analytics

The six Vs of big data

- Big data can be defined by the following characteristics:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.






You may be familiar with the "3 Vs", which were identified first. Later on, more characteristics were added

https://searchdatamanagement.techtarget.com/definition/big-data

What data science can't do

Machines cannot:

- Understand context
- Think through a problem
- Ask the right questions
- Select the right tools
- Interpret results

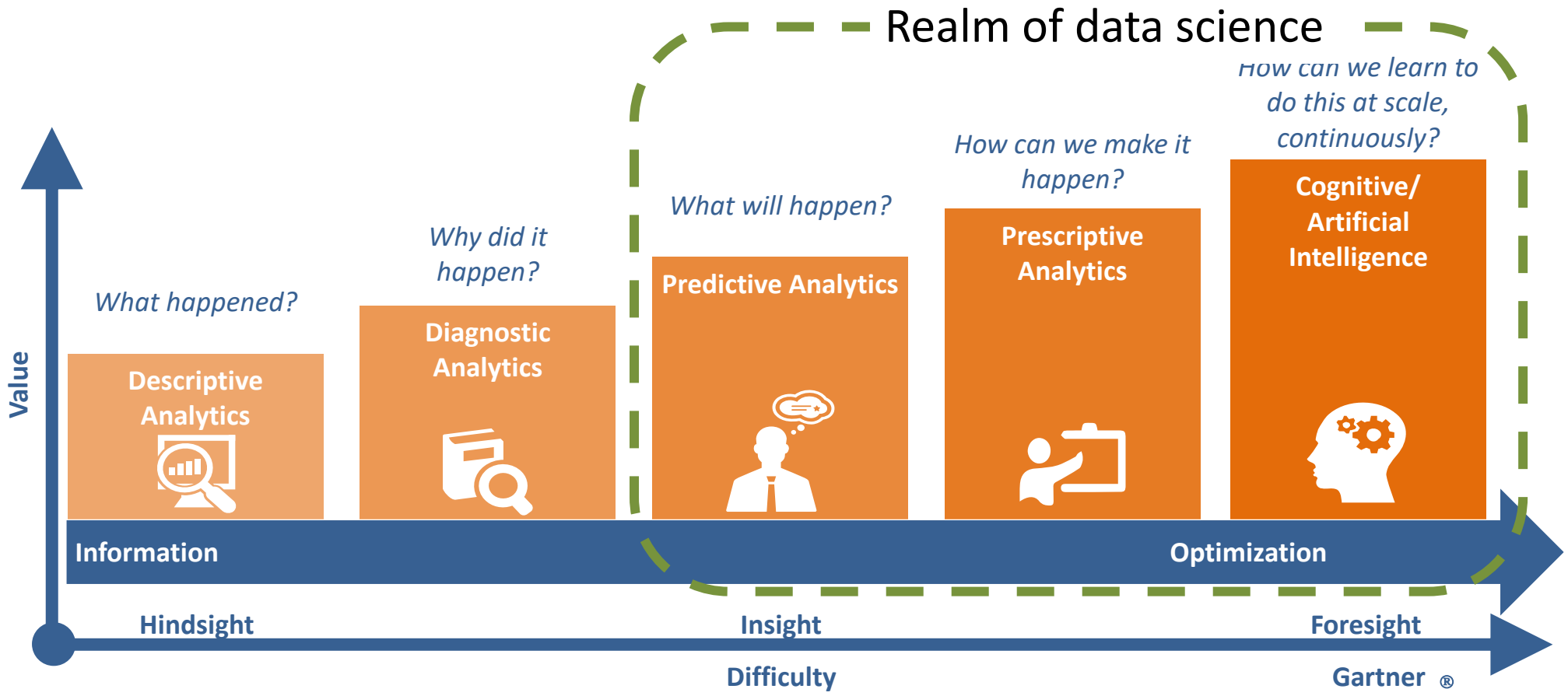
But you can program machines to:

- Perform calculations quickly
- Automate repetitive tasks
- Follow pre-defined rules
- Visualize data

Data science tools are only useful when you have people who can use them effectively

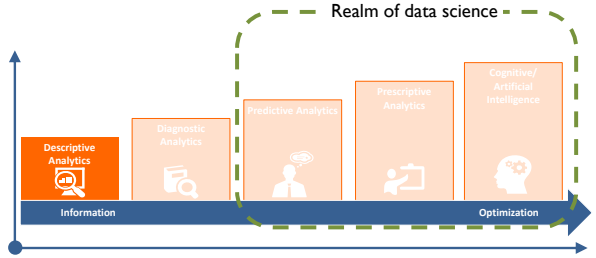
Stages of data analytics

- It takes a coordinated information governance and analytic maturity for organizations to move up the maturity scale



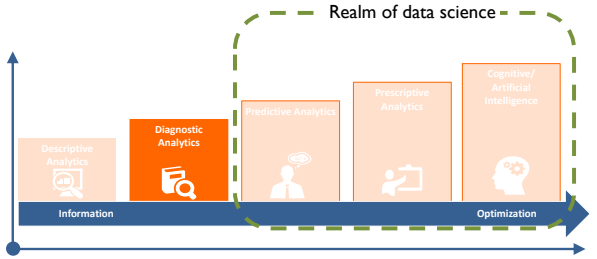
Stages of data analytics: descriptive

Descriptive analytics	
What questions does it answer?	*What has happened in the past?*
How valuable is it?	*Provides some value, but doesn't provide causation or prediction*
How labor intensive is it?	*Easy to deploy, provided that you have the right data*
What tools / methodologies does it include?	*Data dashboards, some BI tools*



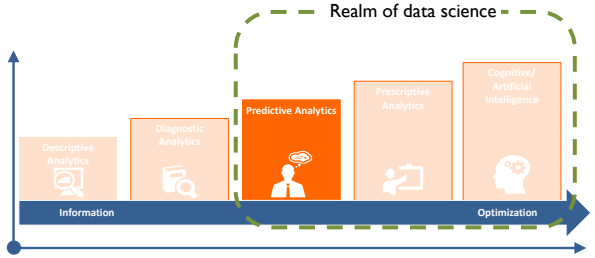
Stages of data analytics: diagnostic

Diagnostic analytics	
What questions does it answer?	*Why did something happen in the past?*
How valuable is it?	*Provides insights into a particular problem, and can help you identify some root causes for past trends and behaviors*
How labor intensive is it?	*Requires detailed data, but doesn't have to be overly intensive*
What tools / methodologies does it include?	*Data dashboards, some BI tools*



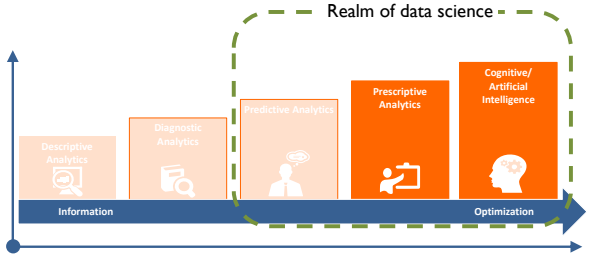
Stages of data analytics: predictive

Predictive analytics	
What questions does it answer?	*What is likely to happen?*
How valuable is it?	*Provides trends / behaviors that are likely to happen*
How labor intensive is it?	*Requires detailed data, and may require a moderate to high level of computer power, depending on the method and the amount of data*
What tools / methodologies does it include?	*Some BI tools, supervised machine learning (regression / classification)*

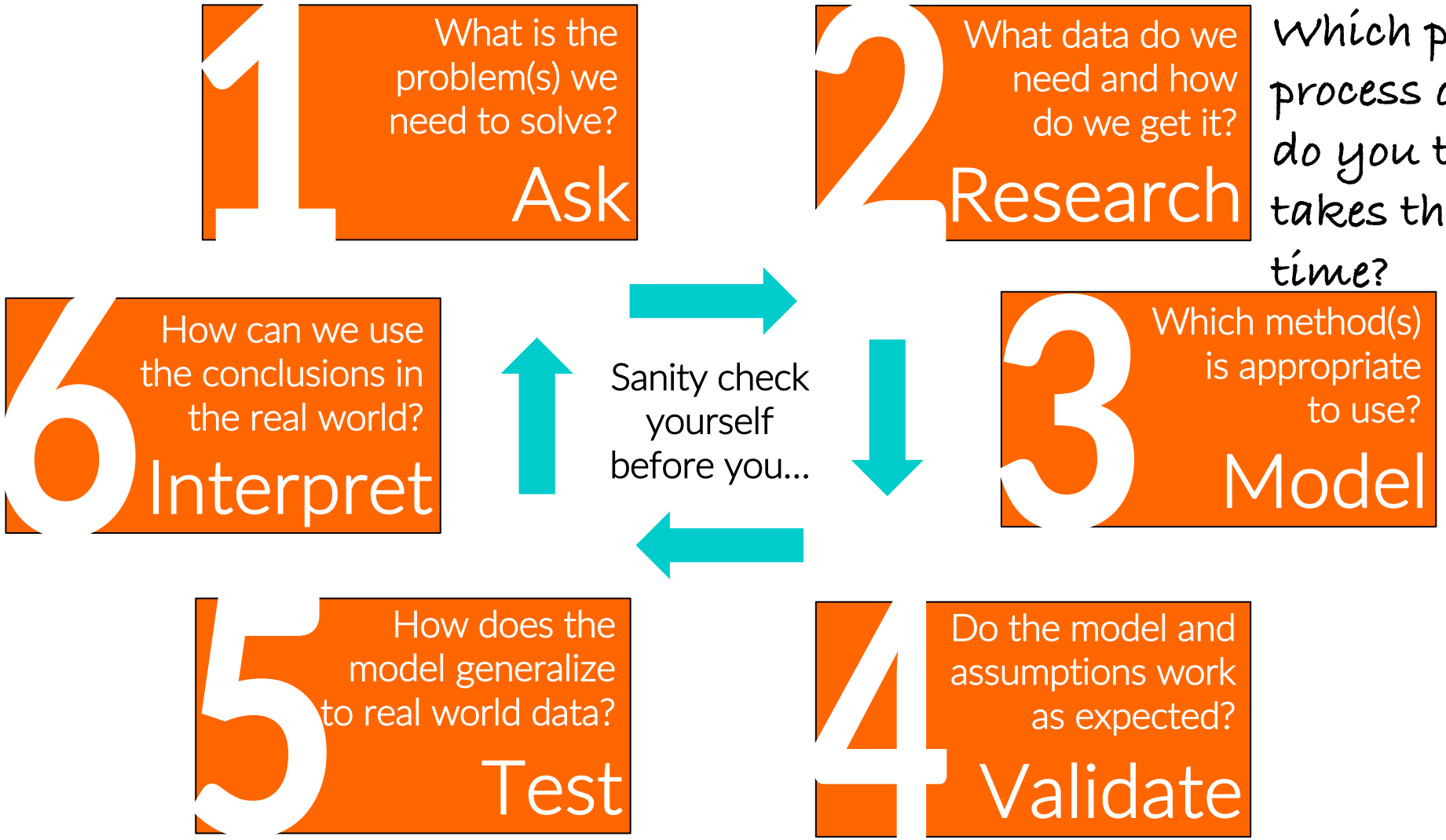


Stages of data analytics: prescriptive / AI

Prescriptive analytics	
What questions does it answer?	*What action should I take next?*
How valuable is it?	*Provides recommendations for future actions*
How labor intensive is it?	*Requires a lot of detailed data, as well as data from other external sources that will impact the model; very labor intensive*
What tools / methodologies does it include?	*Supervised machine learning (regression / classification), neural networks and deep learning*

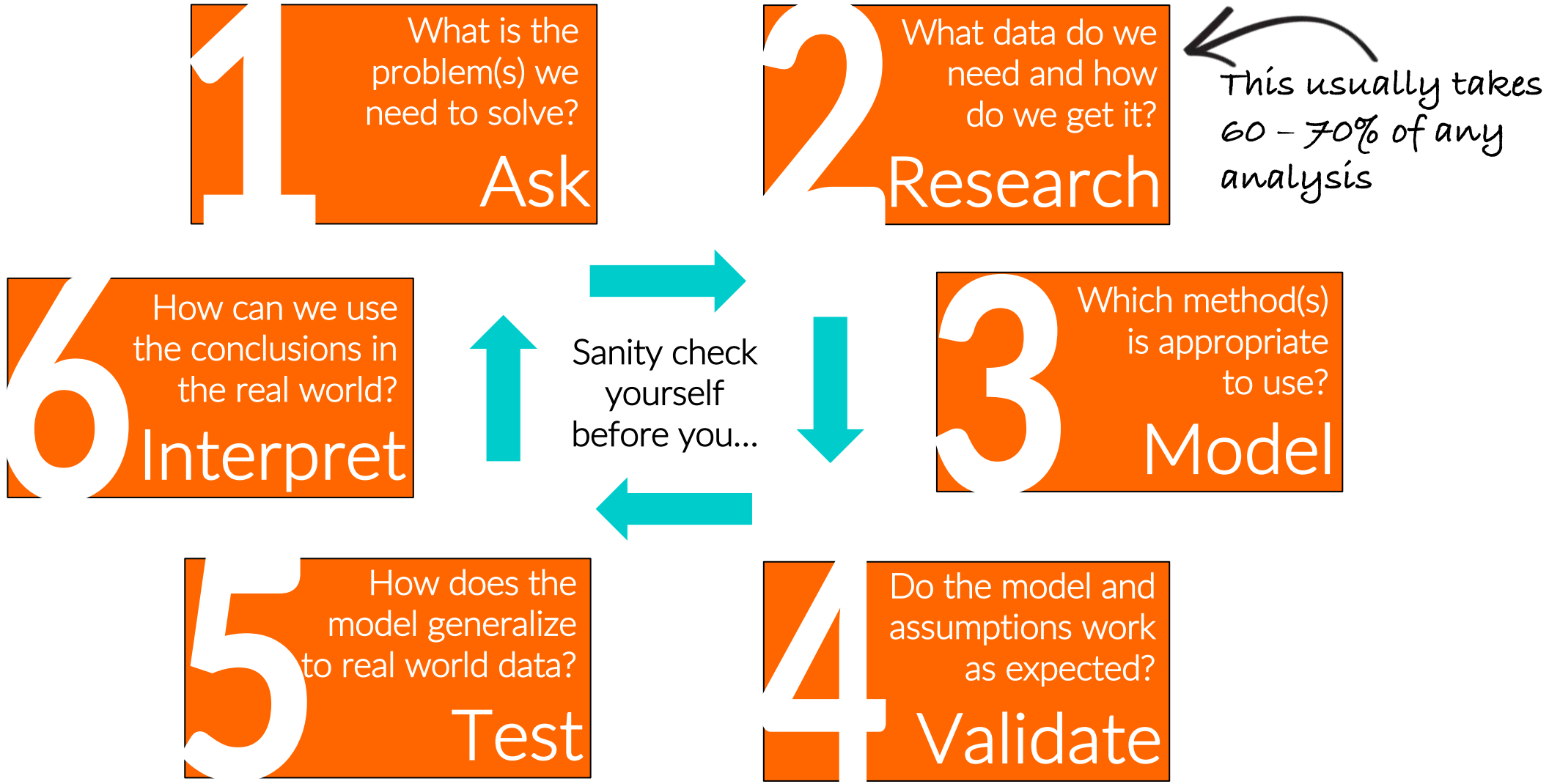


Typical data science process

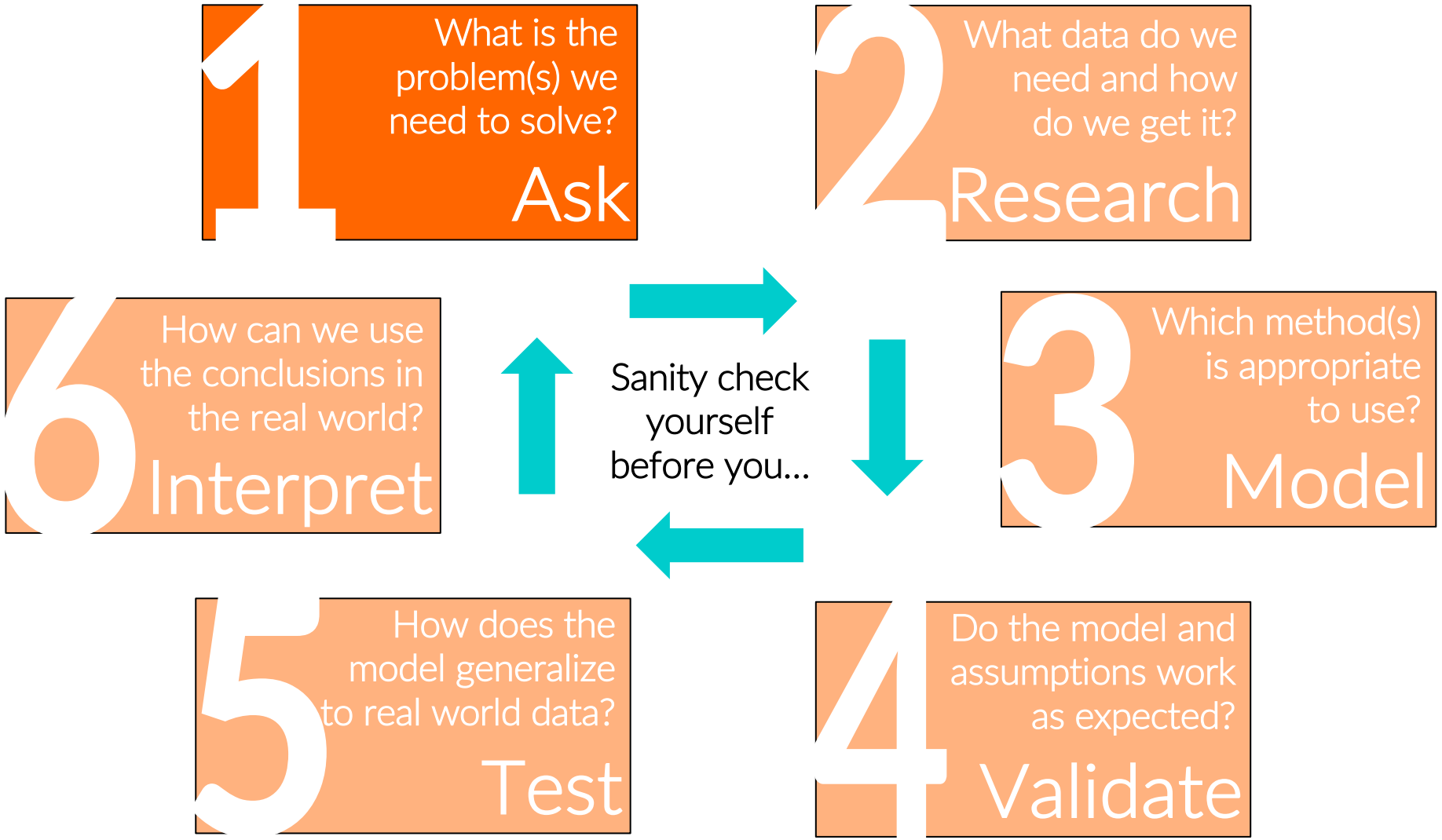


Which part of the process or analysis do you think takes the most time?

Typical data science process



Typical data science process



Asking the right question

- As a leader, one of your most important responsibilities is asking the right questions
- Let's review the framework to set you on a path to data success!

1. What do you want to find out?

- What are your main mission objectives, and what outcome do you want in order to consider an analysis successful?
- Be specific!

How can I make my policies more effective?

Which 3 policies have demonstrated the best results and did they have anything in common?

2. What metrics will you use?

- Once you've identified your question, what are the standard metrics you'll use to measure the results?
- For the question we identified in the previous slide, how do we define 'best results'?

We'll use an indicator that shows the most improvement

We'll use the calculated ROI, and the percent difference in desired behaviors from before and after

3. What data and tools will you use?

- Now that you've defined your question and your metrics, you'll need to make sure that you have the appropriate dataset
- The best datasets have individual-level data as opposed to aggregate data

I'm sure we have the data somewhere

We'll use the datasets from the policy reports that can be found in *x* repository

4. Is your data reliable and high quality?

- Data quality relies on *accuracy*, *completeness*, *timeliness*, *and consistency*
- How has the data been collected, and is there additional documentation?

I'm sure the data is good enough as is

Where can I read about how the data was collected, and the metrics definitions?

5. What data science techniques will you use?

- Now that we have the question, the metrics, and the data, which data science techniques do we want to use (we'll go more into depth in the next section)?
- Typically, you should always perform *exploratory data analysis* beforehand to get an overview of your data

Since I only know one technique, I'll use it for every question

Which techniques are best used to answer the question that we're asking with the data that we have?

6. Who are the stakeholders for this analysis?

- The end users for the analysis will help you decide how to write up the results
- What do the end users need to know from the results and what is their technical background?

I'll put the results in the same report format as I usually do

How can I best convey the results that matter most to my end users?

7. How should you visualize your results?

- Your visualizations need to be accurate, clear, and convey the importance of your results
- Make sure that you're conveying the data accurately, *not the results that the stakeholders are hoping to see*

I'll put the results in the same visualization format as I usually do

How can I best visualize the results that matter most to my end users?

8. Bonus: is your staff data driven?

- Lead by example – if your colleagues / staff is not data driven, show them how you are using metrics to improve efficiency
- We will go over some ways to develop a data driven culture later in this workshop

I'll use data occasionally and sporadically

By consistently using data to drive decisions, we are improving our operations

Sum up

- Even though you might not be executing all the steps, it's important that you map out the process with your team to streamline the analysis:

1. What do you want to find out? *Specific and measurable*
2. What metrics will you use?
3. What data and tools will you use?
4. Is your data reliable and high quality?
5. What data science techniques will you use?
6. Who are the stakeholders for this analysis?
7. How should you visualize your results?
8. Bonus: is your staff data driven?

Activity: re-form your questions!

- Turn to your participant guide to the **Project brainstorm** page to identify 3- 5 ideas for leveraging data in your workplace
- Write each on a separate Post It note (the questions and ideas should be related to the strategies discussed)
- Then, discuss your ideas with your group, and write them down in the appropriate quadrant based on feasibility and impact

Activity time: 15 - 20 minutes

Outline for today

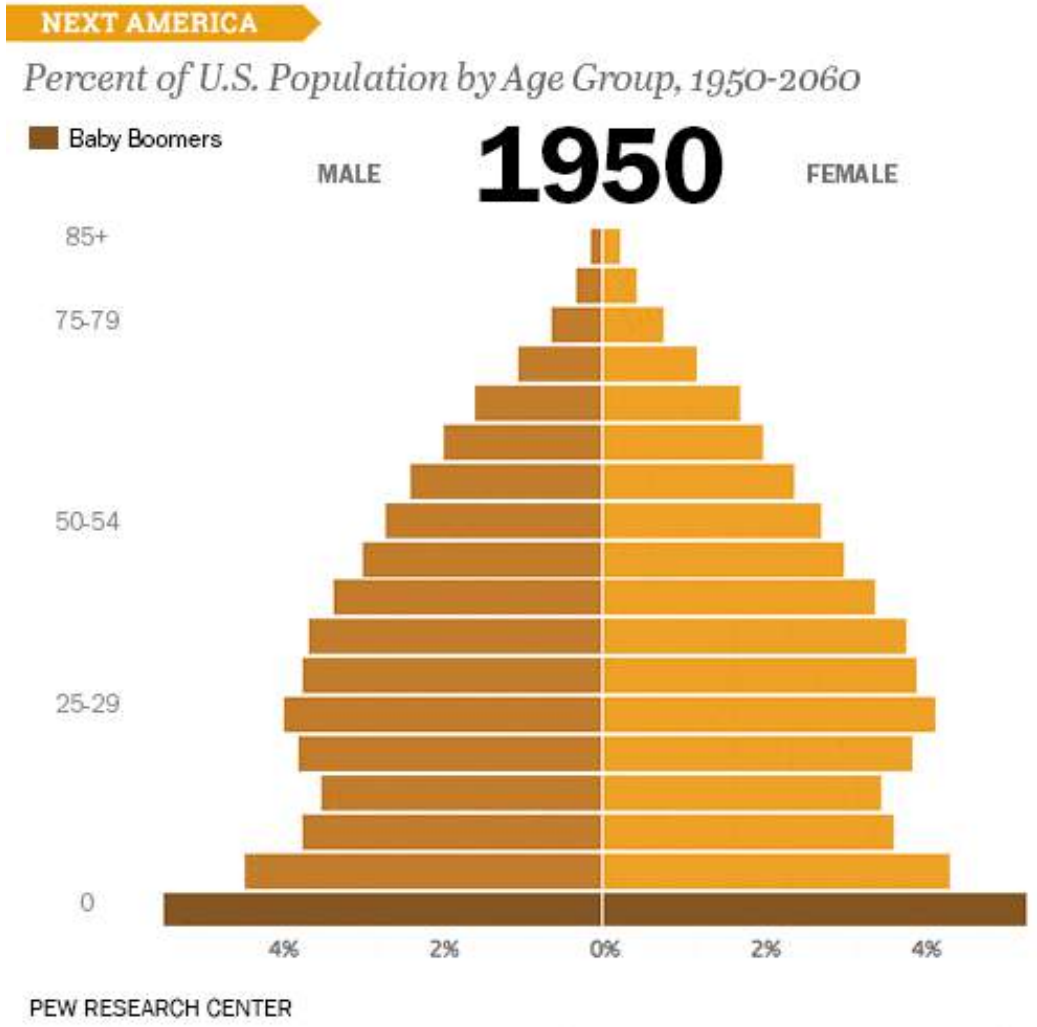
1. Data driven benefits
2. Overview of data science methods
3. Data visualization and storytelling
4. Data ethics frameworks
5. Open data sources

Visualization topics

1. Tools
2. Common graphs and charts
3. Design principles
4. Presentation and storytelling

Why visualize?

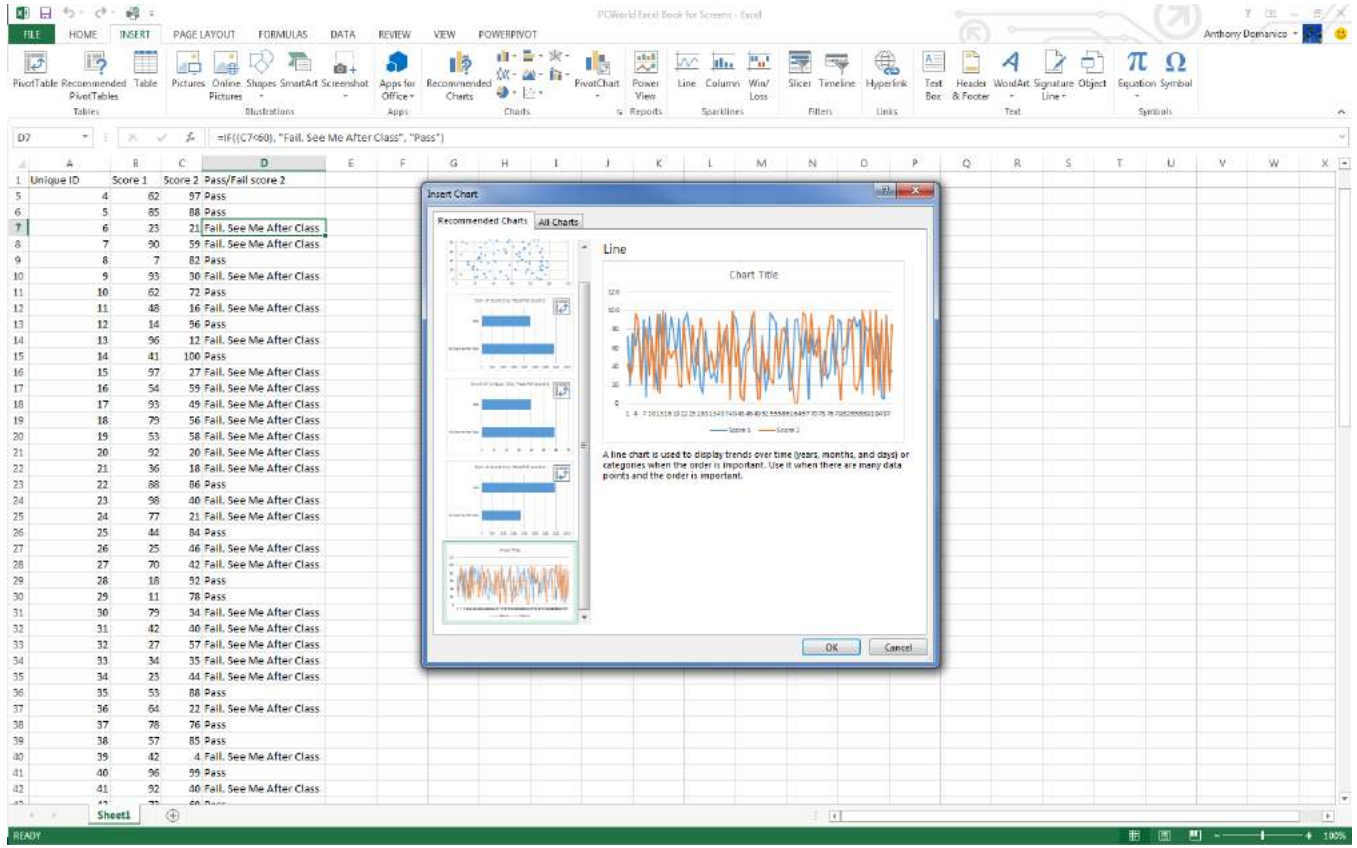
- To provide valuable insights, visualizations must be **interpretable** and **relevant**
- To give a visual or graphical representation of data / concepts
- To communicate ideas - our culture is visual and visualizations are basically visual communication
- Provide an accessible way to see and understand trends, outliers, and patterns in data
- Confirm a hypothesis about the data



Microsoft Excel

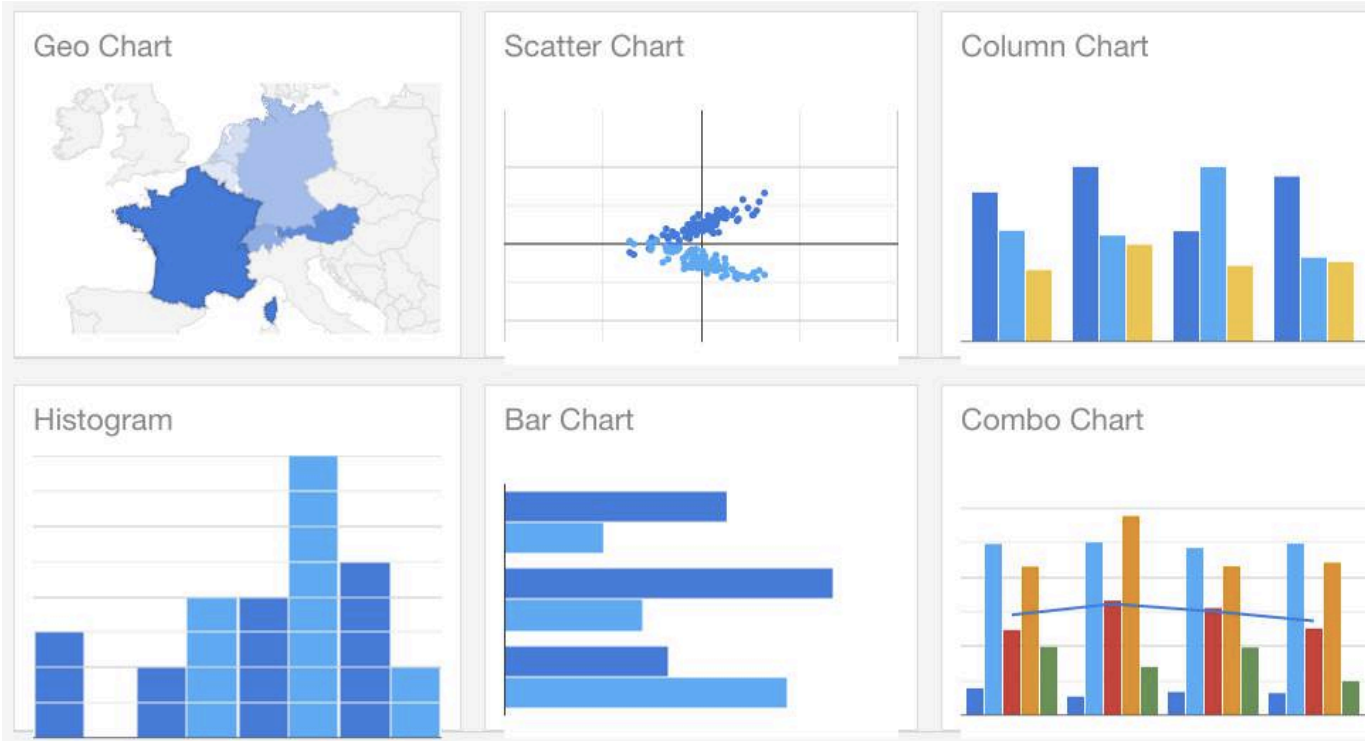
- Create basic chart types such as pie, line, bar, scatter and more

- Charts created in Excel can easily be ported to PowerPoint and Word



Google Charts

- Free and open source which includes a rich gallery, fully customizable, controls and dashboards, and HTML5
- Has more options than Excel; create interactive, animated and geospatial graphics

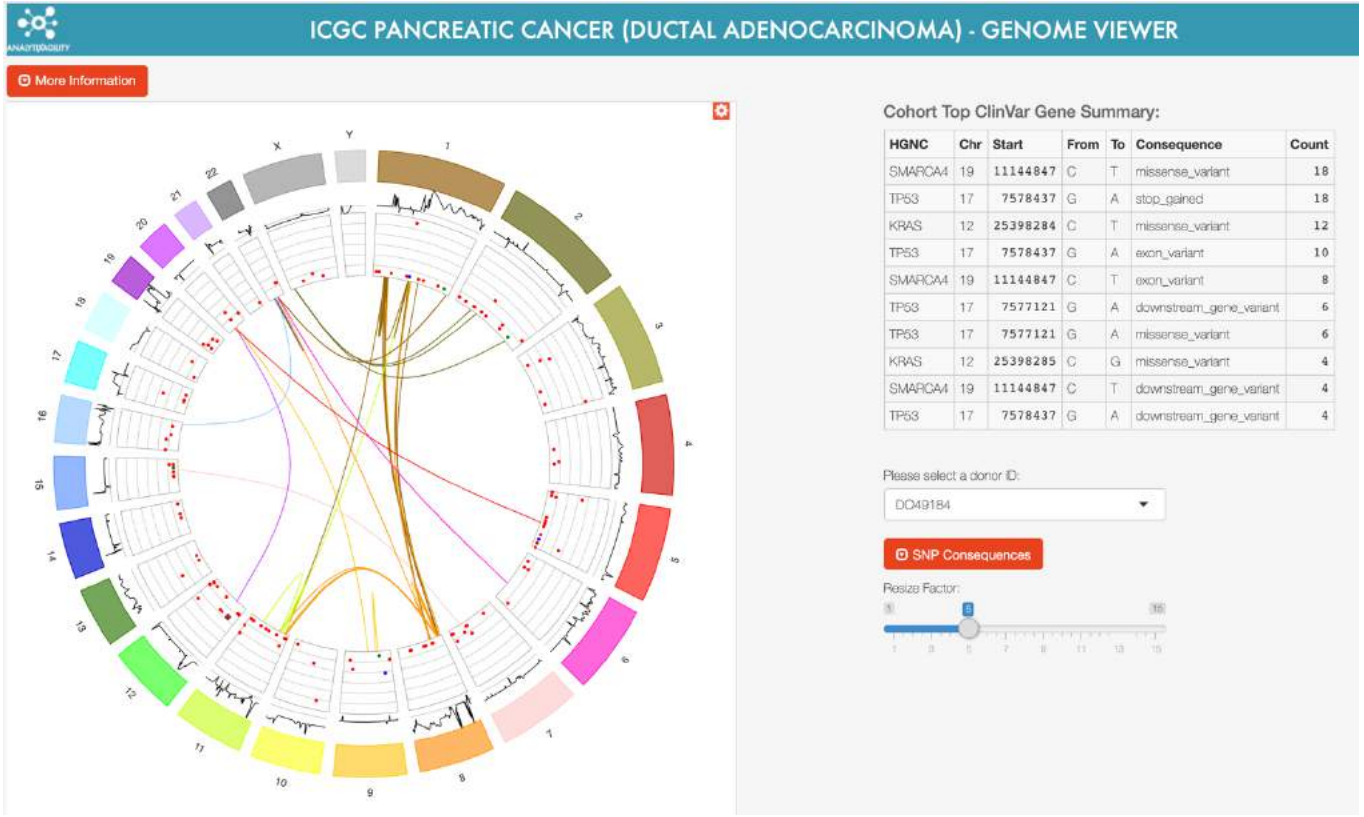


Tableau

- Tool for creating powerful and insightful visuals
- No programming required; drag and drop
- Share and collaborate on premise or in the cloud. Platform can be used department or organization wide

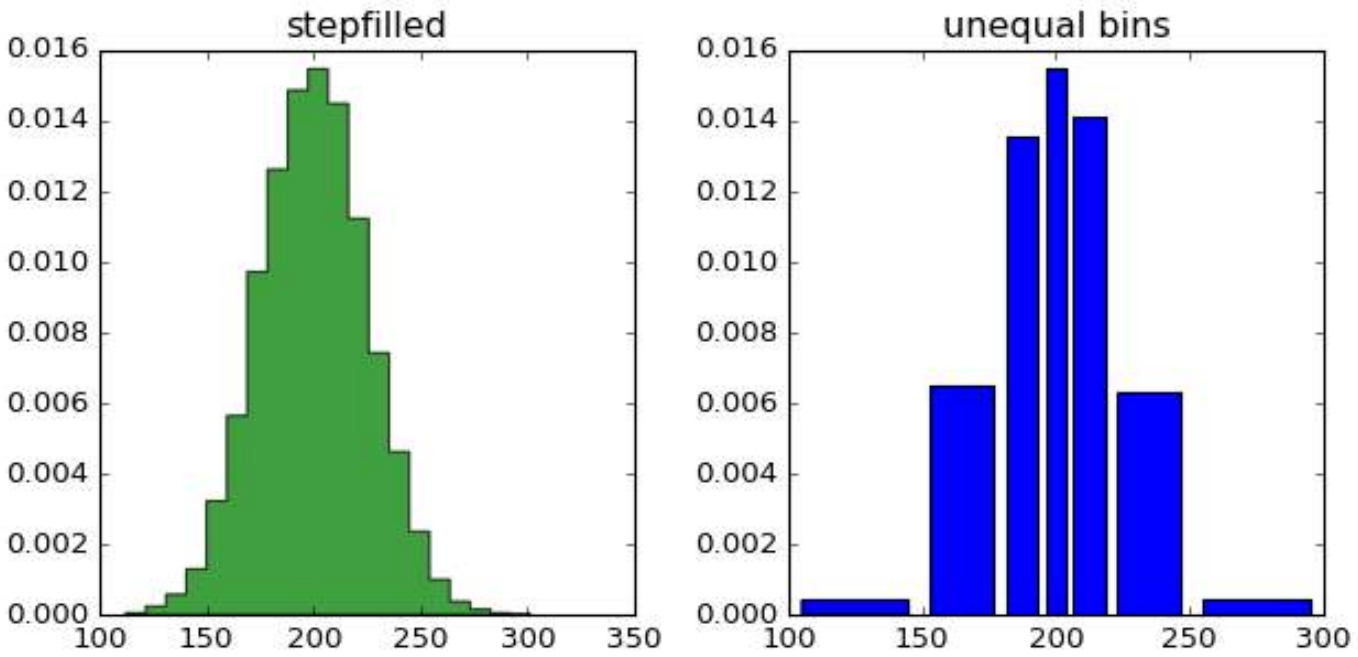
R and RStudio

- Programming tool
- Mainly used for statistical analysis, but has sophisticated packages (code contributed by users) to create interactive dashboards as well
- Open source and free



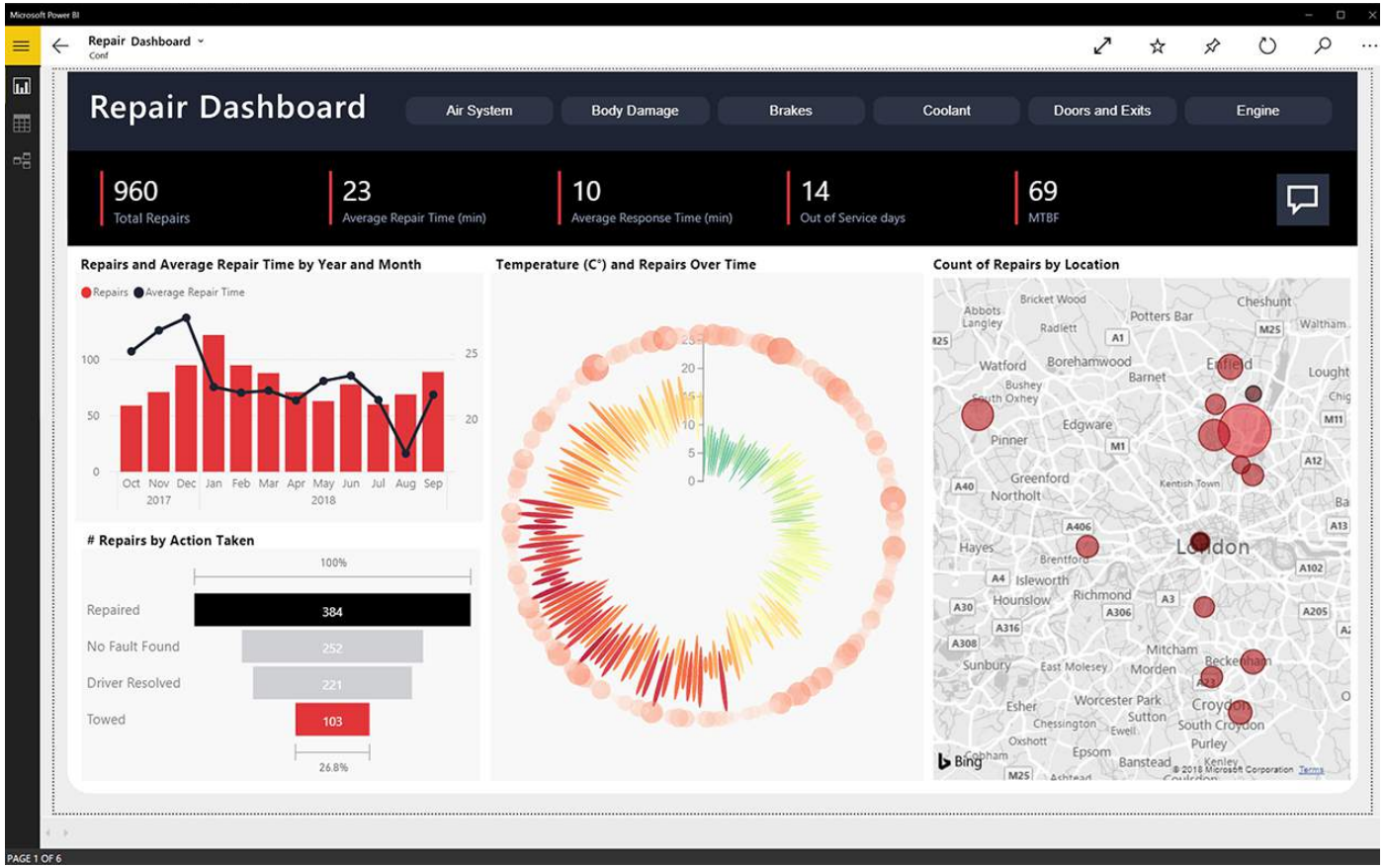
Python

- Programming tool
- Python packages- you'll find libraries for practically every data visualization need
- Matplotlib, ggplot, bokeh, pygal, geoplotlib are examples
- Free and and open source



Power BI

- Interactive visualizations and business intelligence capabilities
- Simple interface
- Create and dashboards

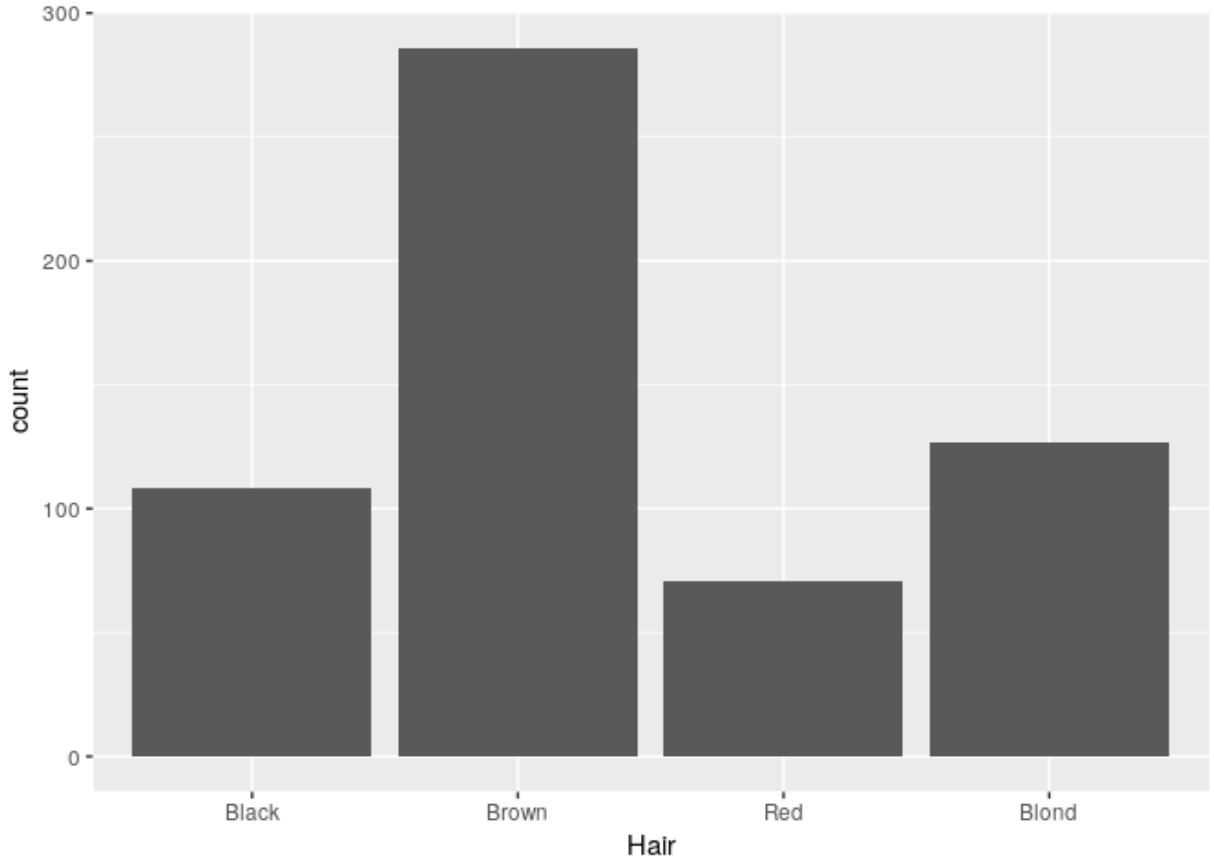


Visualization topics

1. Tools
2. Common graphs and charts
3. Design principles
4. Presentation and storytelling

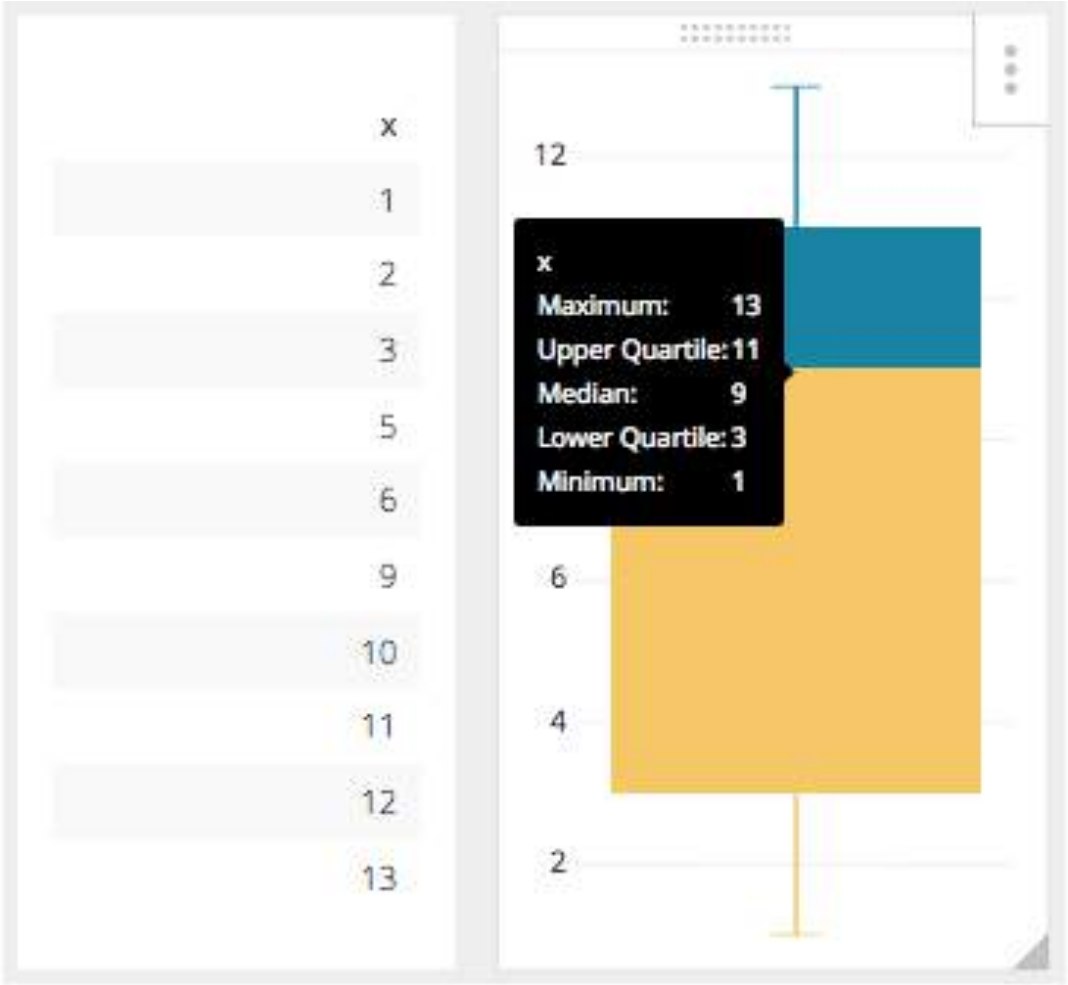
Which visualization should you use?

- Insight - comparisons and proportions
  - Chart types include vertical bar, column bar, horizontal bar, pie, bullet charts, stacked bar, and tree maps
  - Categorical data (i.e. non-numeric or qualitative data)



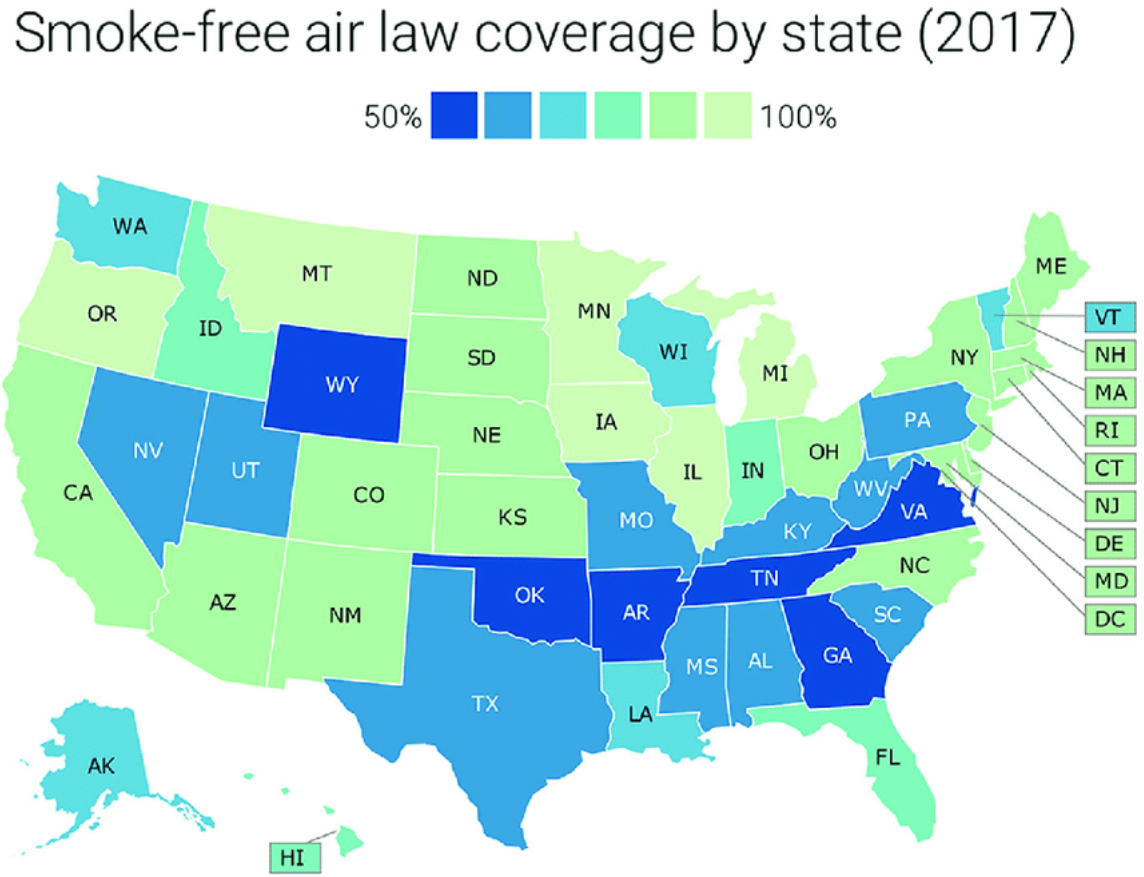
Which visualization should you use?

- Insight - distributions, proportions, frequencies
  - Chart types include histogram, density, and box plots
  - Univariate data (one numeric variable)



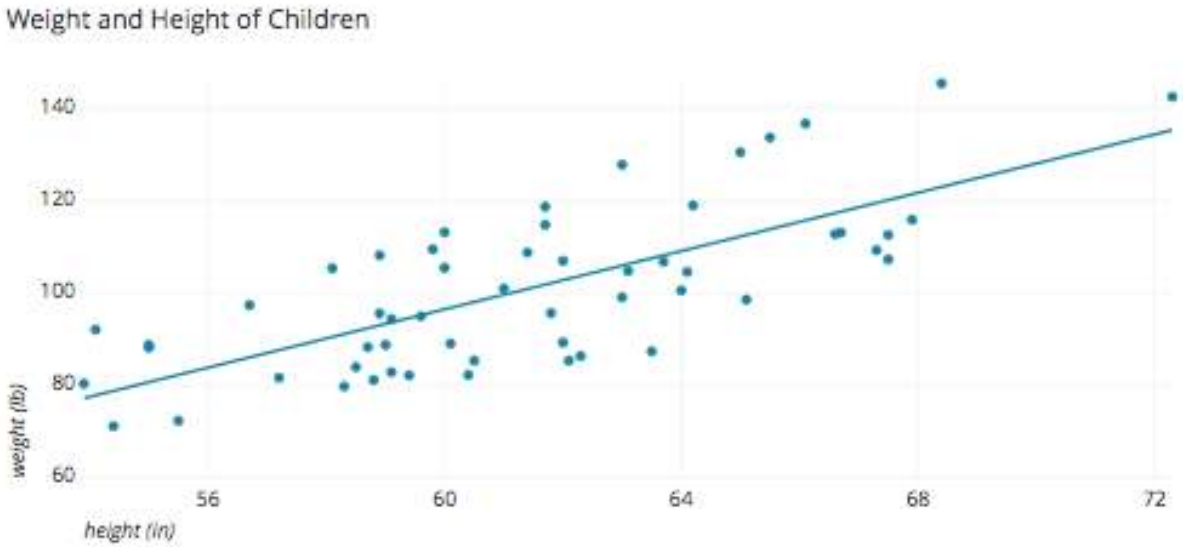
Which visualization should you use?

- Insight - locations, comparisons, trends
  - Chart types include choropleth filled map, point map, connection map, and isopleth map
  - Geospatial data (specific locations)



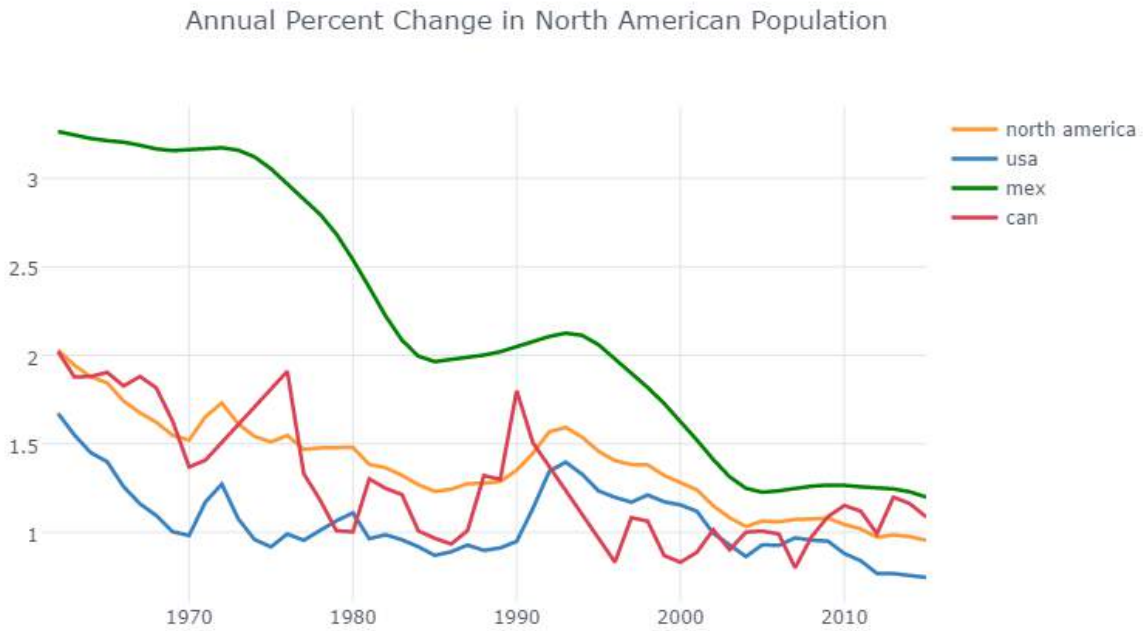
Which visualization should you use?

- Insight – relationships, correlation, proportions, and frequencies
  - Chart types include scatterplot, bubble, parallel, radar, bullet, and heat
  - Two or more numeric variables (i.e. weight, height, and diabetes)



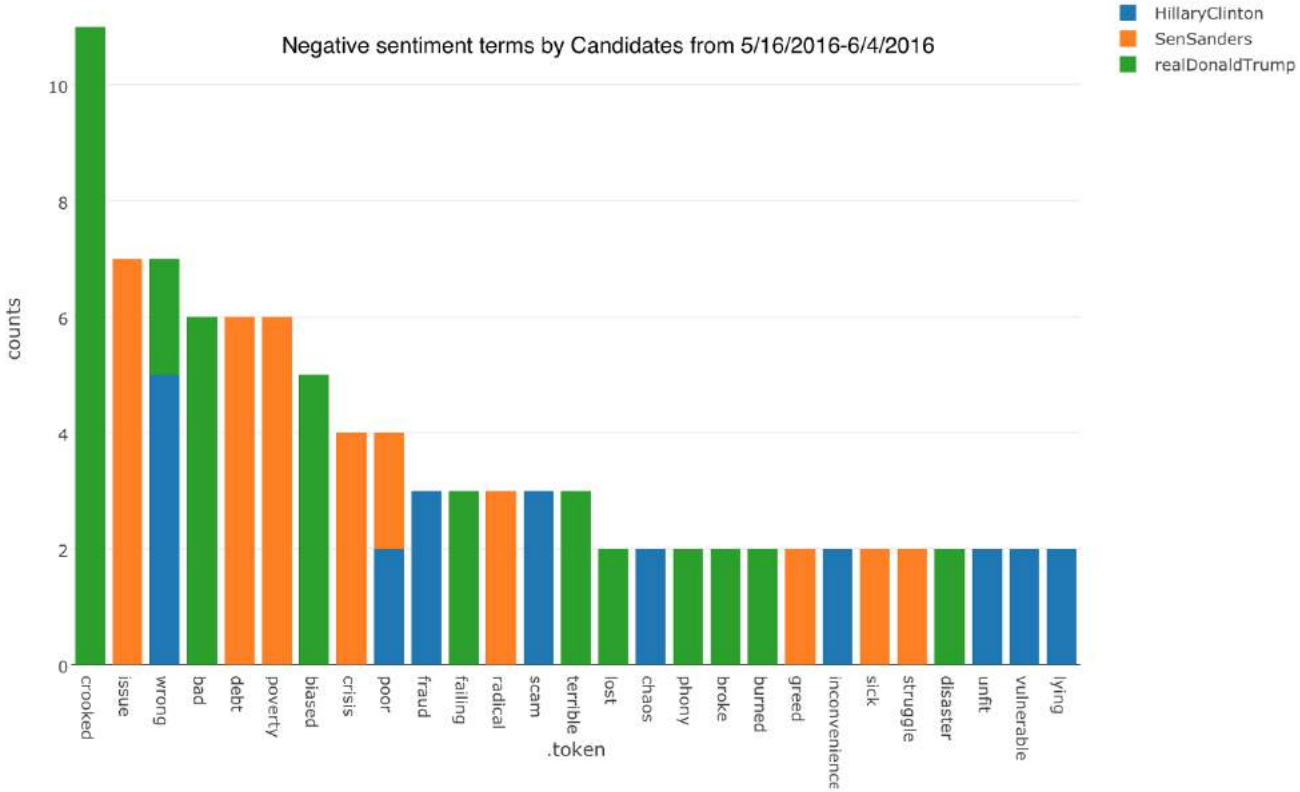
Which visualization should you use?

- Insight - trends, comparisons and cycles
  - Chart types include line, area, bubble, and vertical bar charts
  - Years, months, days, hours, etc.



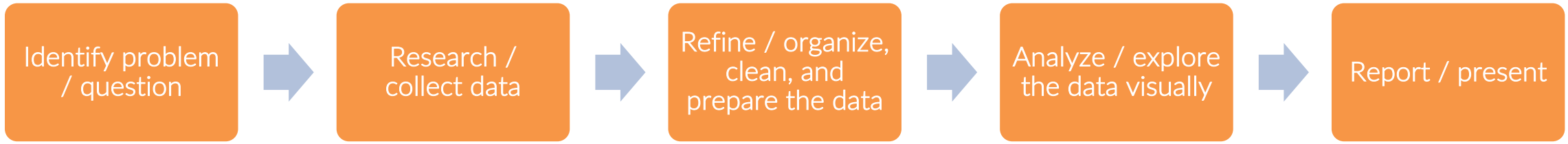
Which visualization should you use?

- Insight - sentiment, comparisons, frequency
  - Chart types word cloud, histogram, bar chart
  - Single words or phrases (keywords)



Data visualization prep and analysis

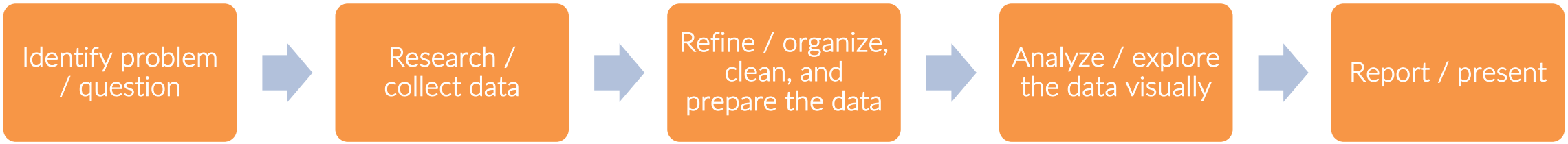
- Before you can start analysis and making visualizations, you should start with a clear problem statement
- Once you have a clear problem statement, you can follow the process below for data analysis and prep, and then explore data visually



Data visualization prep and analysis

Discuss: Think back to your ideas on the sticky notes earlier. Focus on the one that is most feasible

1. What other questions would you need to answer?
2. What data do you think you would need to get started?
3. What visualizations and tools do you think you would use?

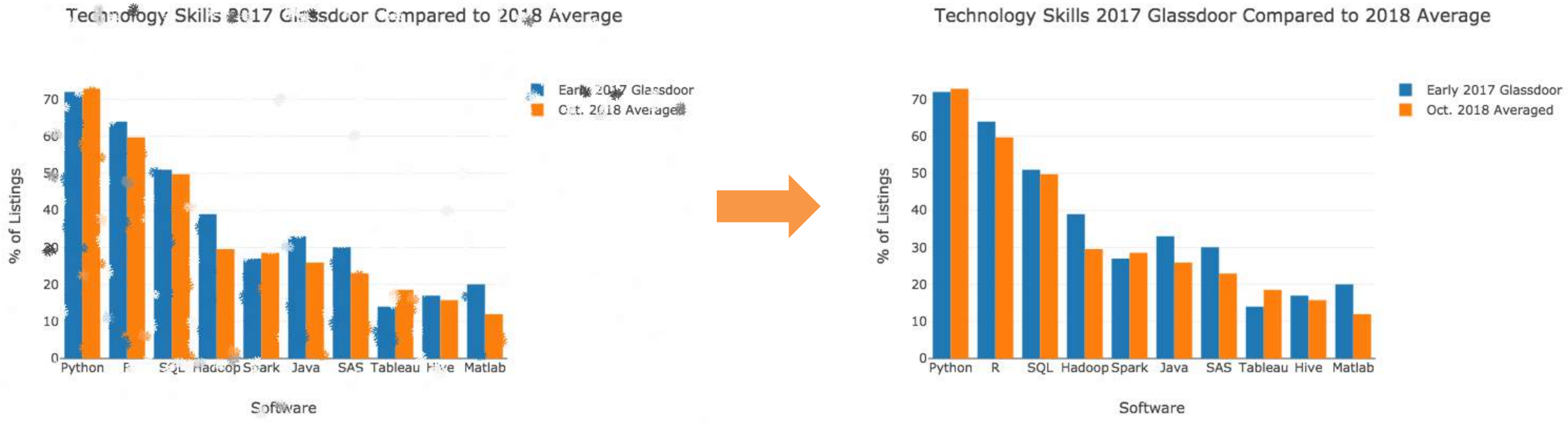


Visualization topics

1. Tools
2. Common graphs and charts
3. Design principles
4. Presentation and storytelling

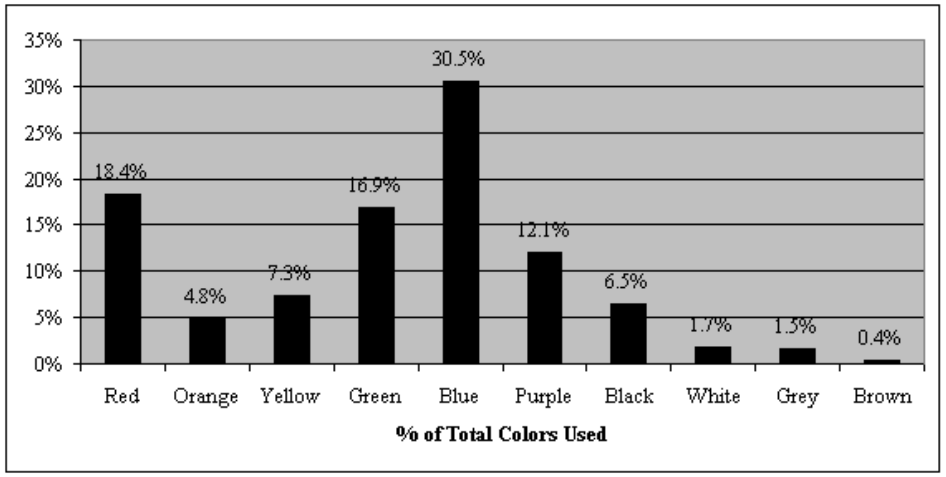
10 design principles

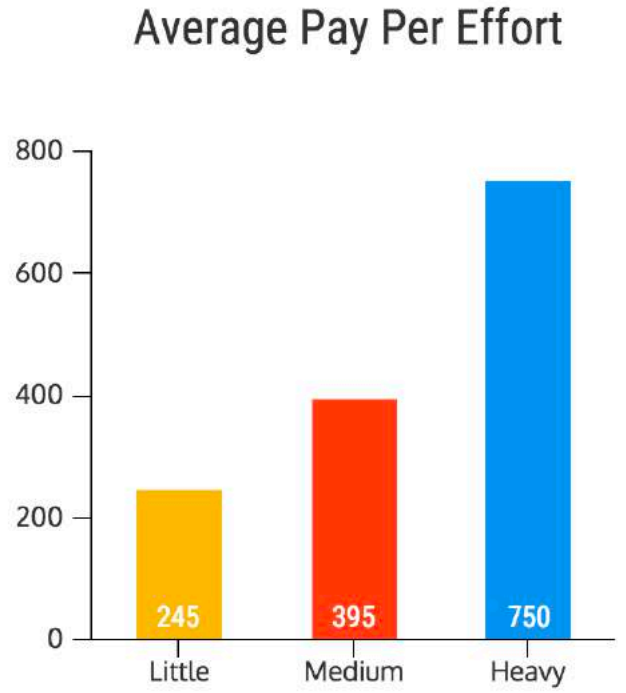
- Chart format - the resolution and file type should be appropriate for where it will be displayed



10 design principles

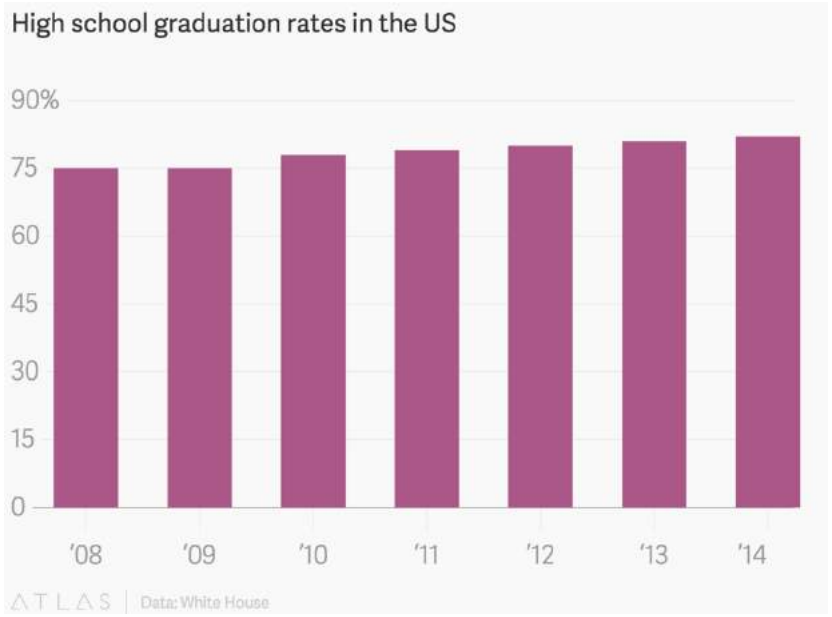
- Color - use color only when it related to differences in the data. Ensure high contrast values to be mindful of color-blindness and other visual sensitivities





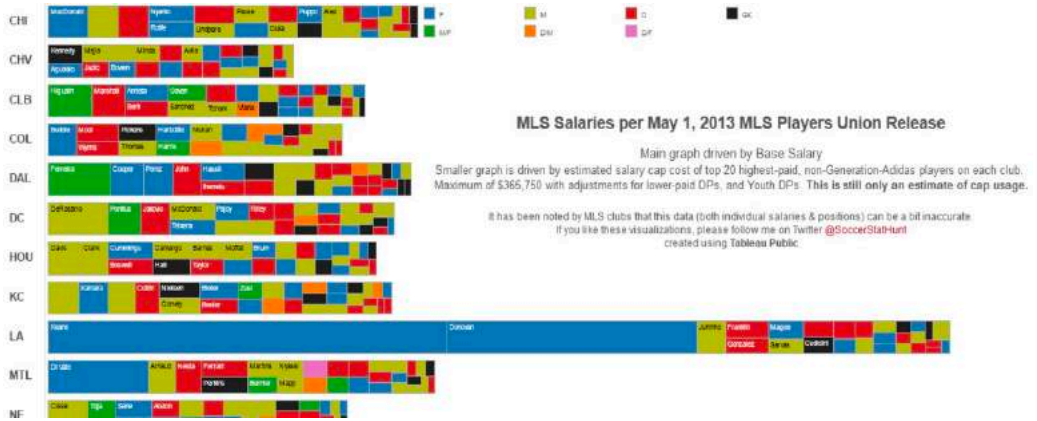
10 design principles

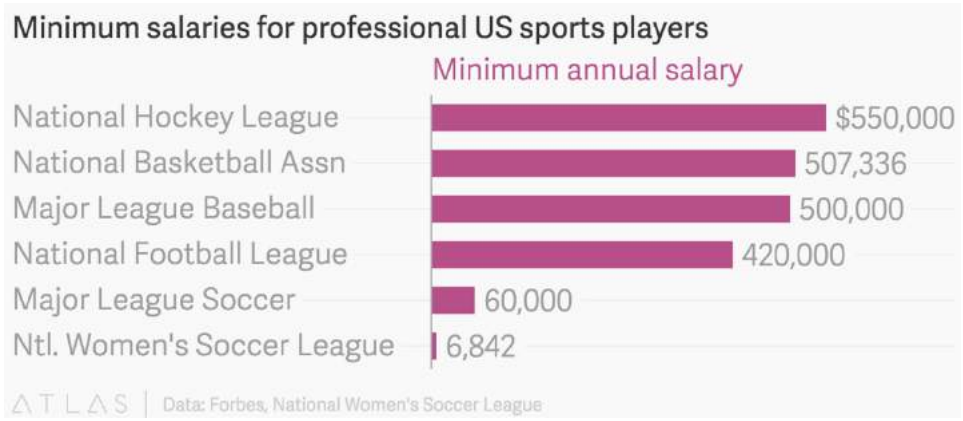
- Text and tables - descriptions can guide your readers, and communicate key insights. Use legends when necessary, make sure to label axes and include a heading



10 design principles

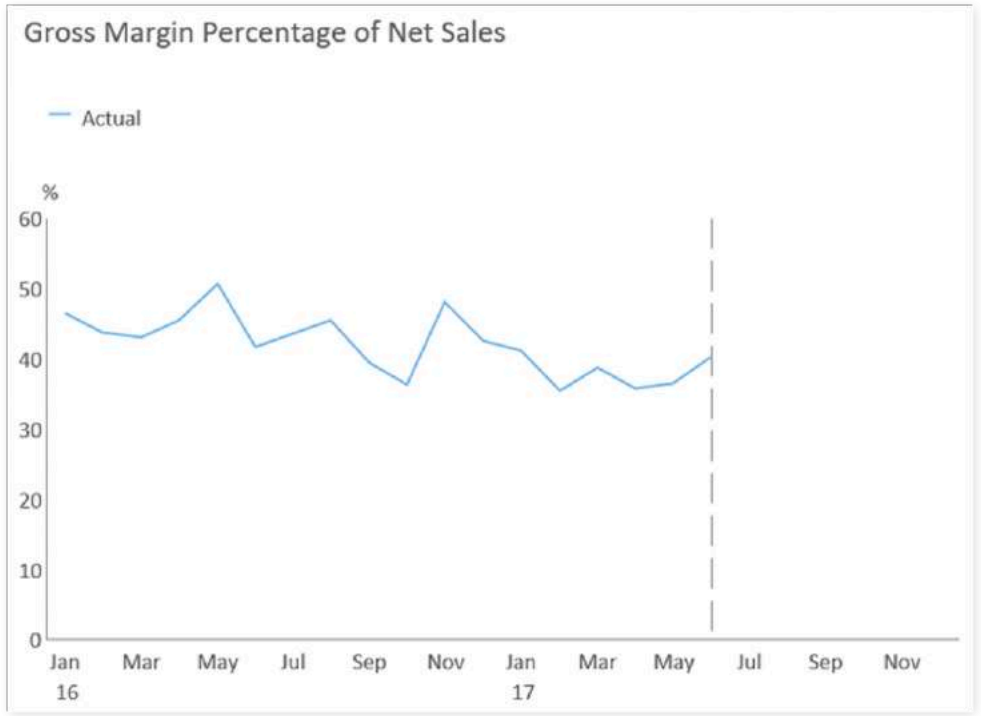
- Readability - ensure everything on your chart is readable

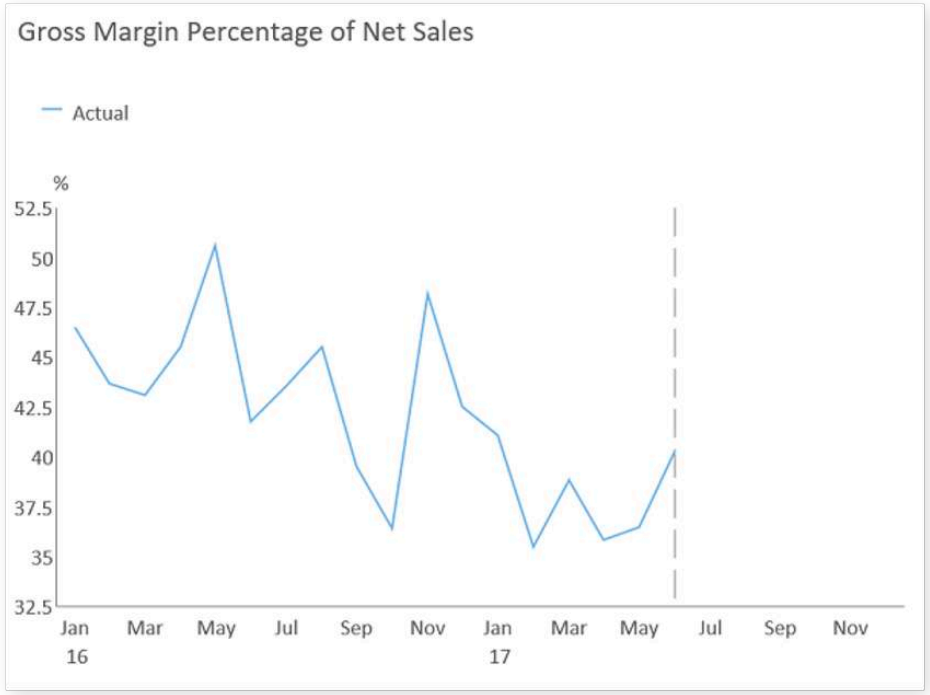




10 design principles

- Scales - use natural increments on your axes (1, 2, 3; 0, 100, 200, 300); make sure that the scales are proportionate and intuitive



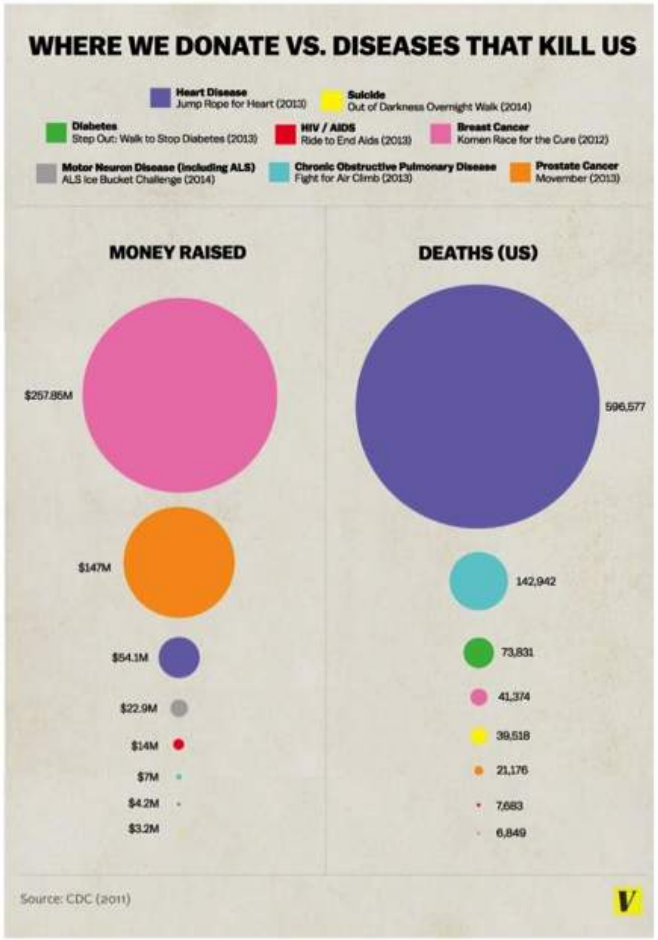


10 design principles

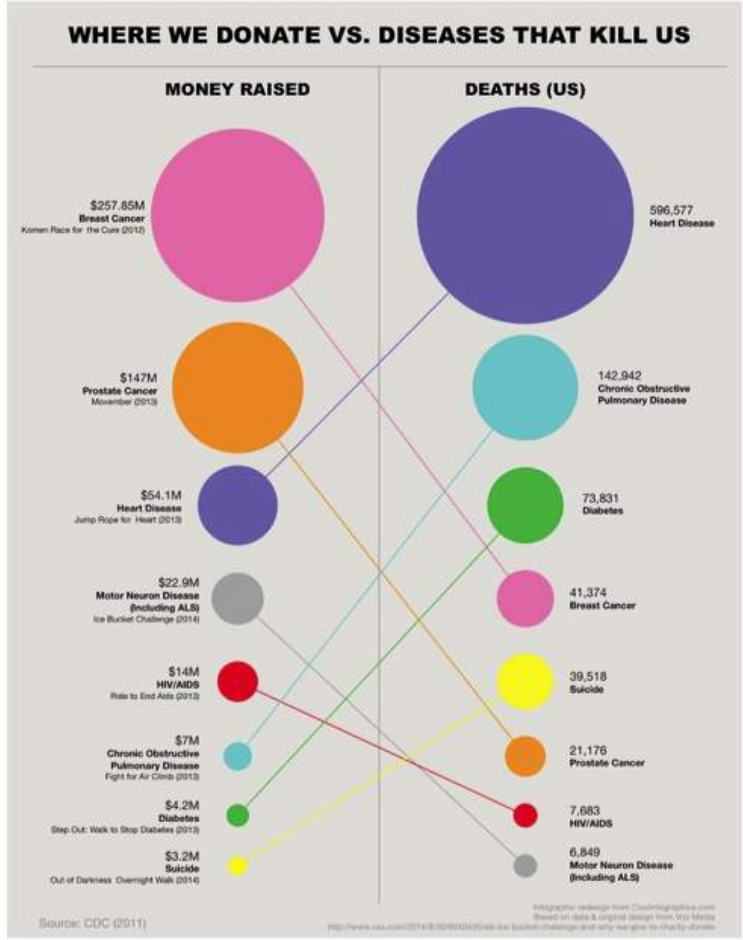
- Data integrity - keep the lie factor equal to 1

- Visualization expert, Edward Tufte, states: *"The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the quantities represented."*

Original Design

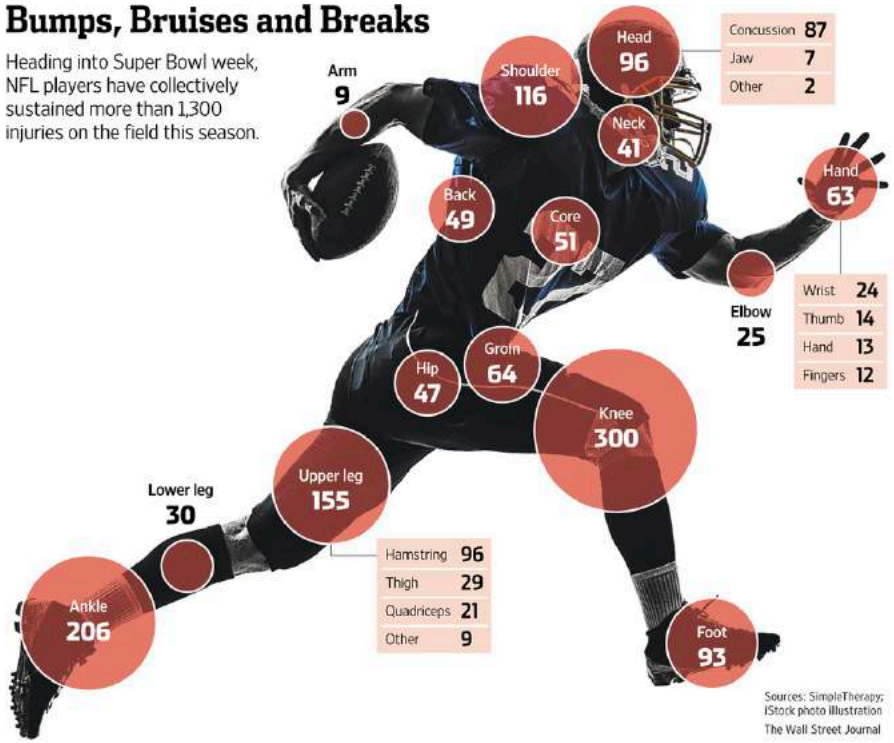


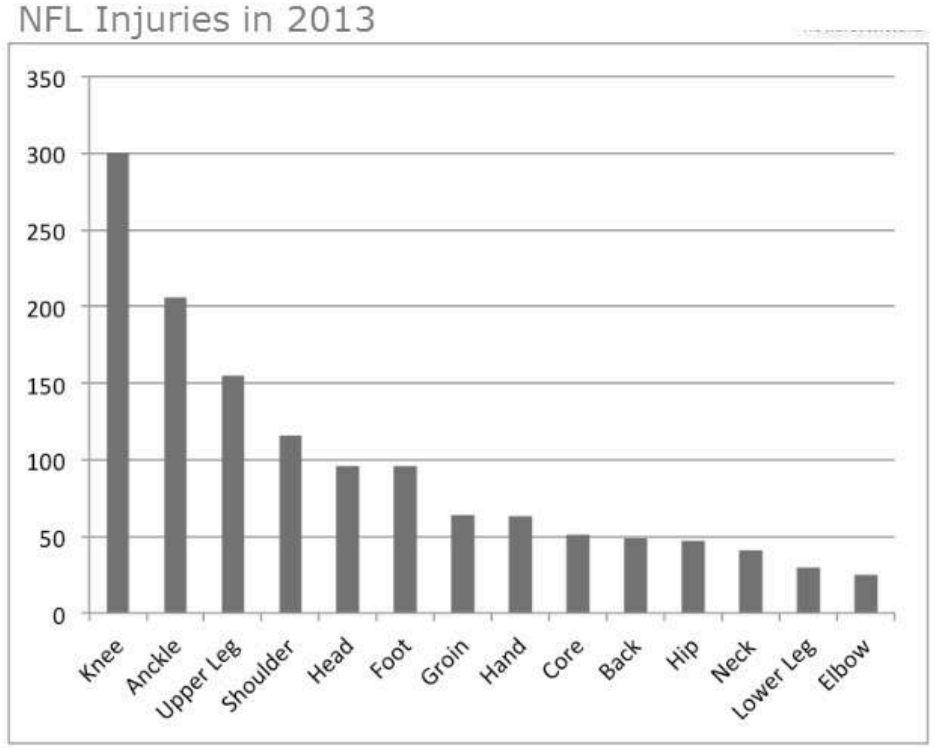
Corrected Design



10 design principles

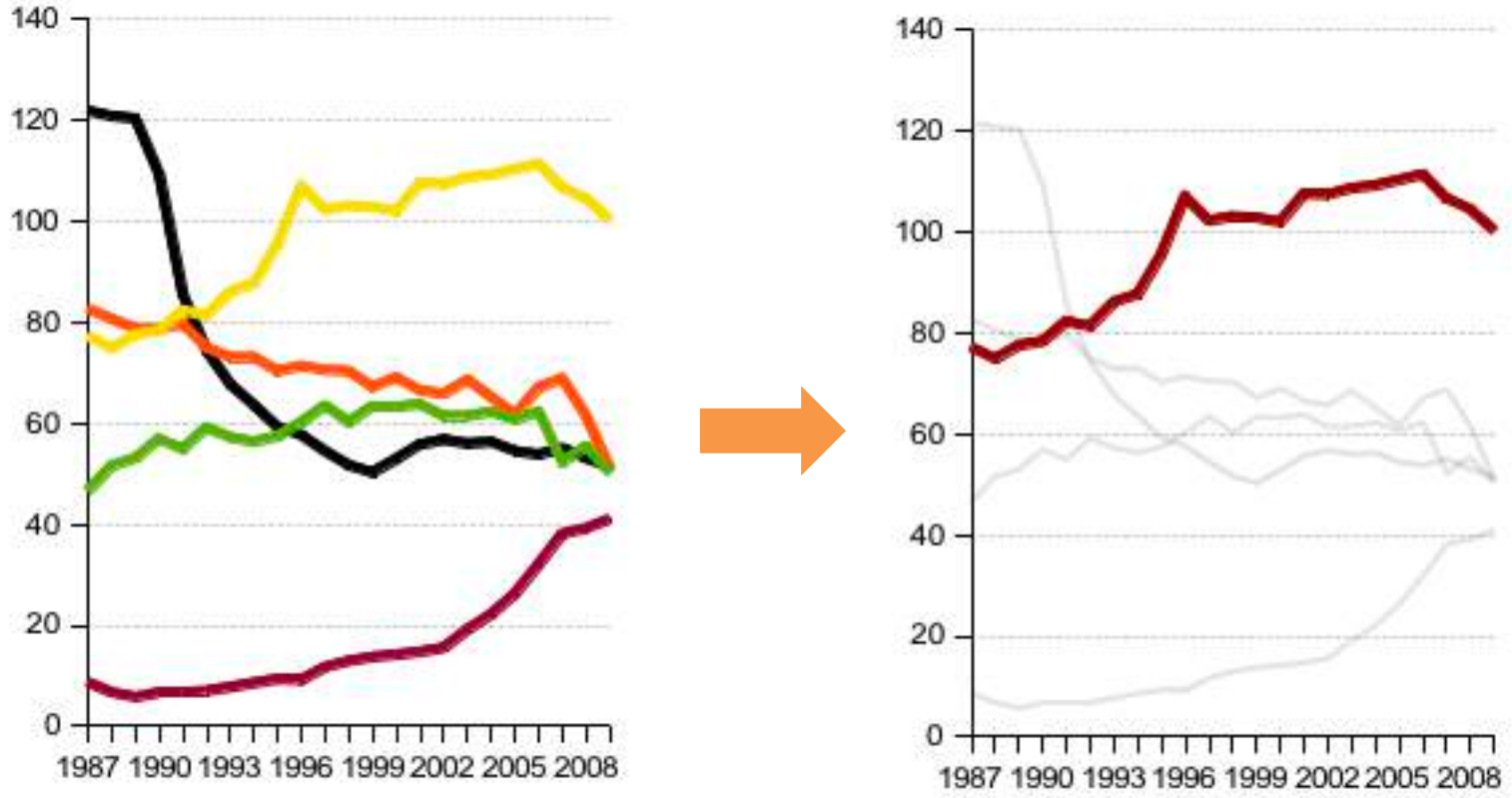
- Chartjunk - remove grid and any other non essential or non data elements





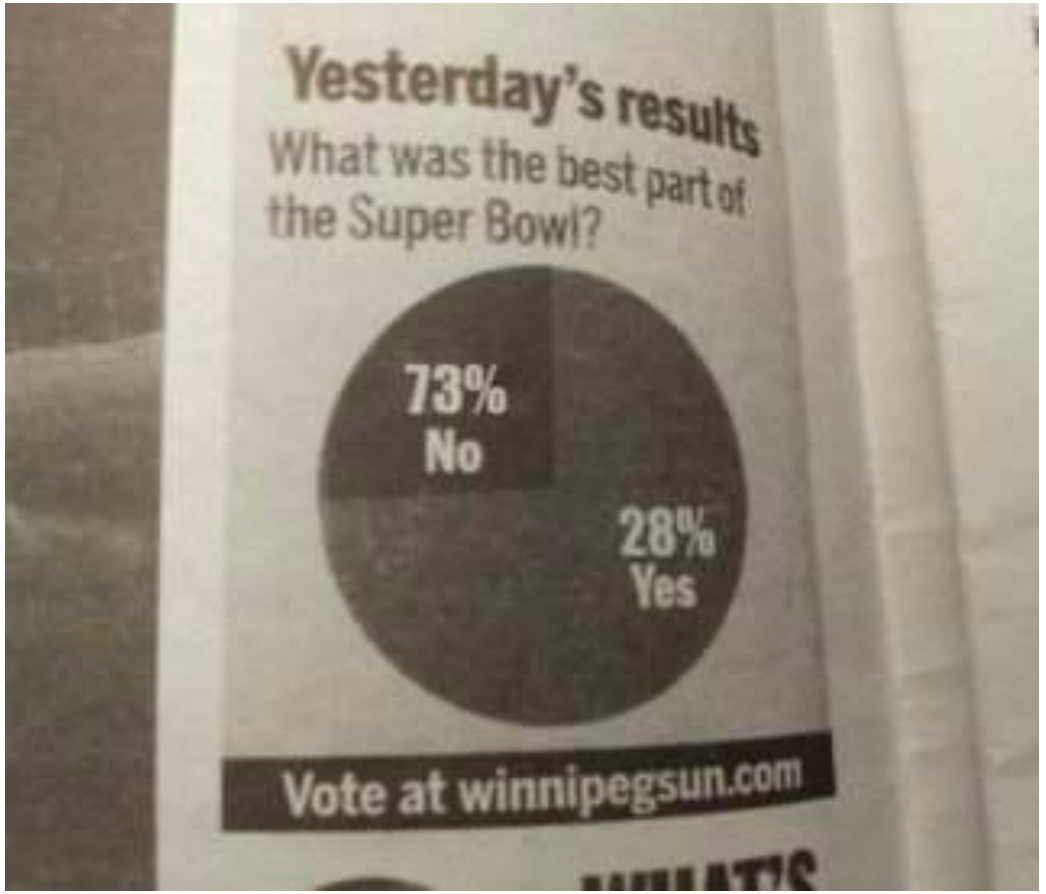
10 design principles

- Data density - don't present too much data on a single chart / graph



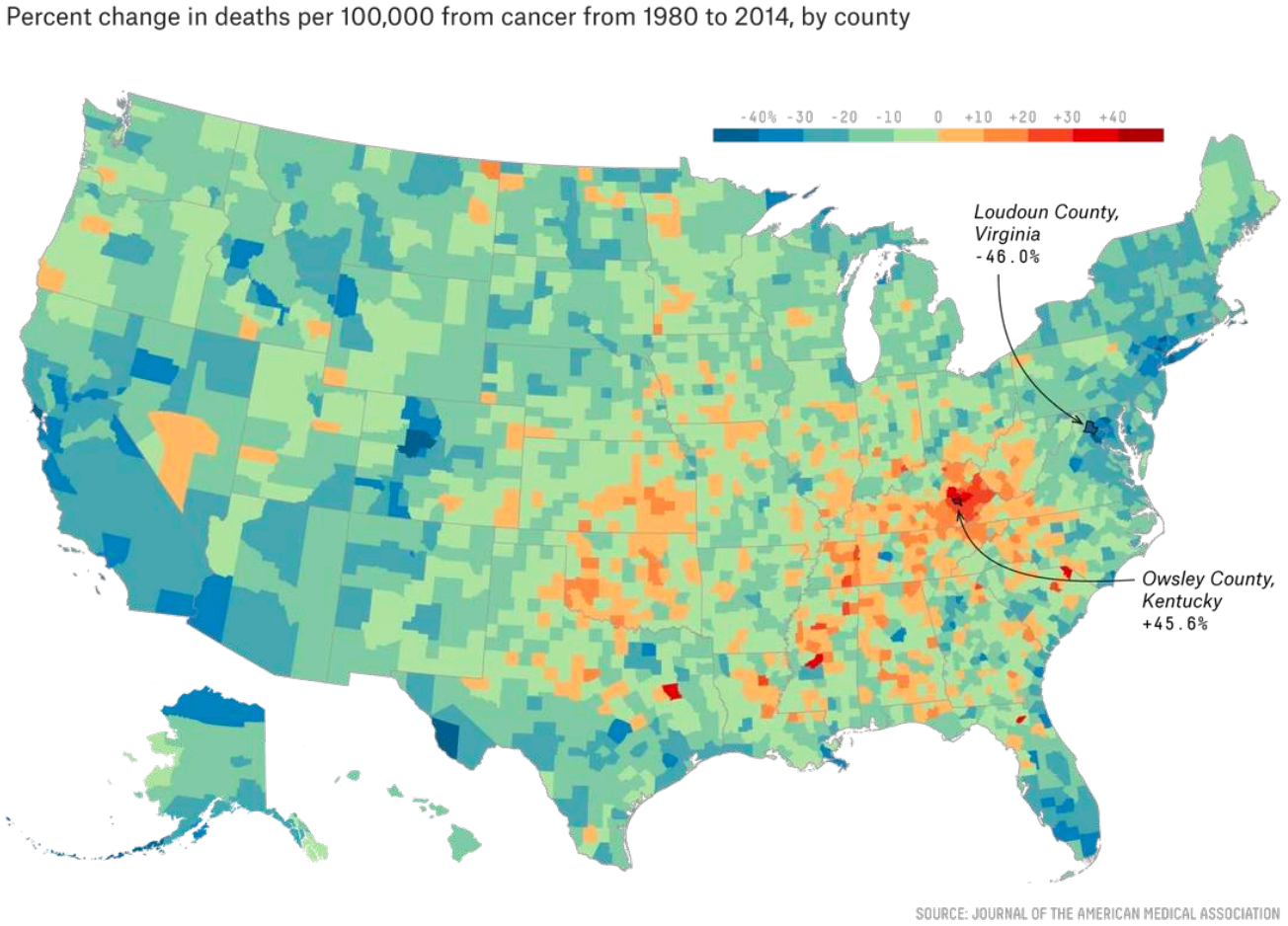
10 design principles

- Data richness – is it high-quality data from reliable sources and accurate?



10 design principles

- Attribution - include a citation to the source of the data in your visualizations



10 design principles summary

- As you're building out your visualization, keep these principles in mind:
  1. Ensure an appropriate resolution / file type 
  2. Use high contrast colors meaningfully 
  3. Use labels and descriptions 
  4. Make sure your chart is readable 
  5. Use accurate and proportionate scales 
  6. Make sure comparative sizes are accurate 
  7. Remove unnecessary information 
  8. Clear out dense charts 
  9. Double check data reliability and accuracy 
  10. Include citations for your data sources 

Activity: assess visualizations

- Turn to your participant guide to the Analyzing visualizations page to assess the 4 visualizations – do they follow the 10 principles?
- Write down your notes on the visualizations
- Then, discuss your thoughts with your group – did you have the same conclusions? Which chart follows the most principles?

Activity time: 15 - 20 minutes

Congratulations!

1. Data driven benefits
2. Overview of data science methods
3. Data visualization

Tomorrow, you'll learn

1. How to tell compelling stories with data
2. How to implement data ethics in data science
3. How to find various sources of data
4. How to manage and hire data science teams