# ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY: ETHICS IN ACTION

**MARCH 25, 4:30 PM**

FOUNDERS GREAT HALL,
HAVERFORD COLLEGE

How does the US Department of State use Artificial Intelligence to support analysis and national security? **What are the ethical concerns involved in that practice, and how are they mitigated?** How is AI, as a tool, enhancing security and safety while policymakers and administrators make critical decisions about how and where it will be deployed? **What vital decisions will be considered over the years to come?**

Join a Haverfordian conversation with leading experts in this area.

Please, no press and no recording

## Speakers

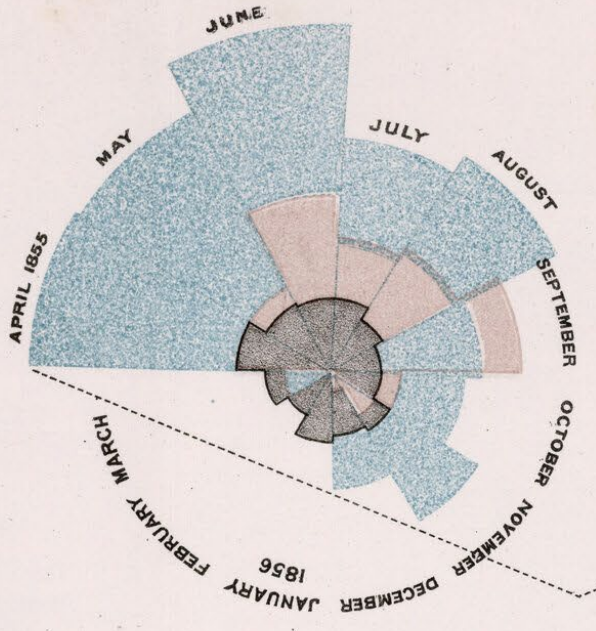
**JOHN SILSON '02,**
DIRECTOR OF THE US DEPARTMENT OF STATE CENTER FOR ANALYTICS

**SORELLE FRIEDLER**
SHIBULAL FAMILY PROFESSOR OF COMPUTER SCIENCE

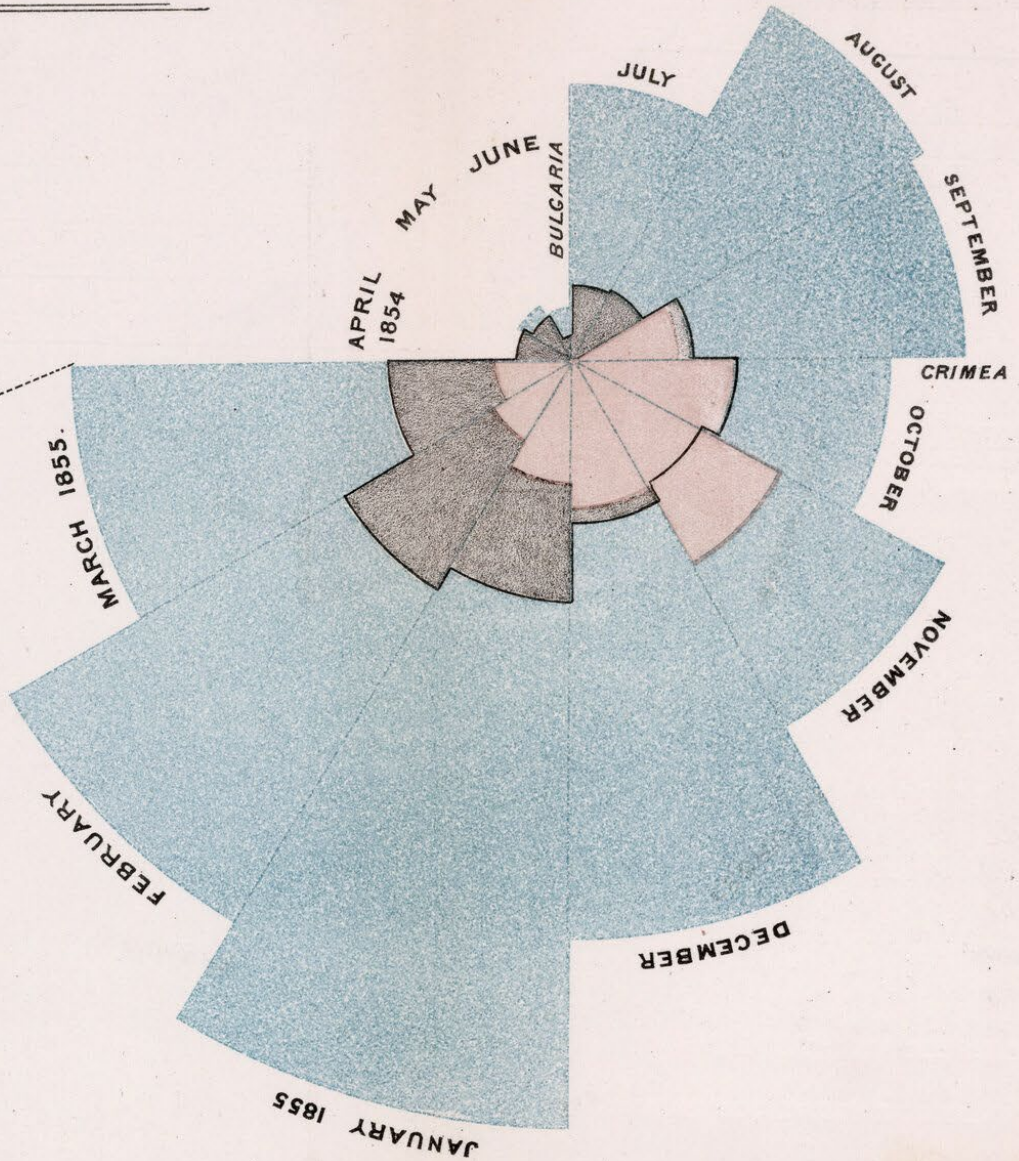
# Visualizations

# DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.




1.
APRIL 1854 TO MARCH 1855.




*The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.*

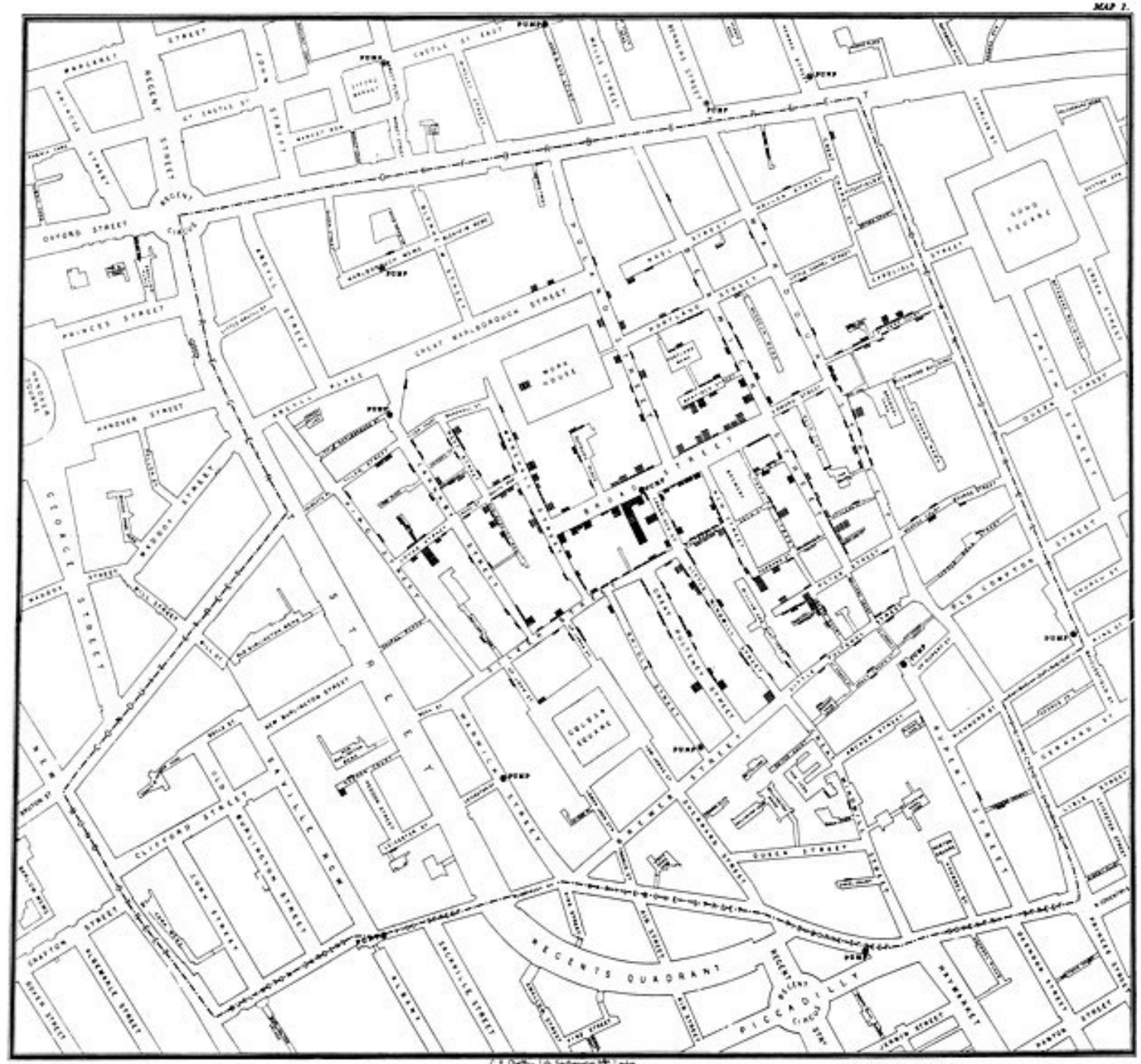
*The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.*

*The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.*

*In October 1854, & April 1855, the black area coincides with the red; in January & February 1855, the blue coincides with the black.*

*The entire areas may be compared by following the blue, the red & the black lines enclosing them.*

Harrison & Sons, St. Martin's Lane.


MAP 1.
OXFORD STREET
REGENT STREET
PRINCES STREET
HANOVER STREET
GEORGE STREET
BROAD STREET
GOLDEN SQUARE
REGENT'S QUADRANT
PICCADILLY
SOHO SQUARE
PUMP
SCALE 80 INCHES TO A MILE.

# Types of Data

*Remember!*
*Data in the same column need to be the same type and comparable.*

| | NUMERICAL | CATEGORICAL |
|---|---|---|
| Definition | Value from fixed scale | Value from fixed inventory |
| Ordering | Yes | Doesn't have to be |
| Differences | Meaningful | Might not be different |

# Terminology

- Individuals
  - those whose features are recorded
  - individual observations
  - rows
- Variables
  - features or attributes of those individuals
  - these can vary across individuals
  - columns

| firstName | lastName | tel | building | officeNumber | faculty? |
|---|---|---|---|---|---|
| Alyssa | Pivirotto | 7917 | Carpenter | A4 | N |
| Alice | McGrath | 7916 | Carpenter | A5 | N |

# Steps to Create a Visualization

1. Define Goal and Audience
2. Understand and Prepare the Data
3. Choose the Right Visualization Type
4. Select the Right Tool
5. Design & Refine

# Why does this matter?

Let’s think through an example with COVID-19 data. How might you want to visualize this data for (1) the public or (2) policy makers or scientists.

Examples:

- John's Hopkins COVID-19 Dashboard
- Estimation of Excess Deaths Associated with COVID-19

# Steps to Create a Visualization

1. Define Goal and Audience
2. Understand and Prepare the Data
3. Choose the Right Visualization Type
4. Select the Right Tool
5. Design & Refine

**audible_uncleaned.csv** (12.07 MB)

Detail Compact Column

### About this file

This file does not have a description yet.

| name | author | narrator | time | releasedate | language | stars | price |
|---|---|---|---|---|---|---|---|
| 82767 unique values | 48374 unique values | Narratedby:anony... 1%<br>Narratedby:矢島雅... 1%<br>Other (85581) 98% | 2284 unique values | 5058 unique values | English 71%<br>german 9%<br>Other (17310) 20% | Not rated yet 83%<br>5 out of 5 stars1 ra... 4%<br>Other (11333) 13% | 586.00 6%<br>668.00 5%<br>Other (77694) 89% |
| Geronimo Stilton #11 & #12 | Writtenby:GeronimoStilton | Narratedby:BillLobely | 2 hrs and 20 mins | 04-08-08 | English | 5 out of 5 stars34 ratings | 468.00 |
| The Burning Maze | Writtenby:RickRiordan | Narratedby:RobbieDaymond | 13 hrs and 8 mins | 01-05-18 | English | 4.5 out of 5 stars41 ratings | 820.00 |
| The Deep End | Writtenby:JeffKinney | Narratedby:DanRussell | 2 hrs and 3 mins | 06-11-20 | English | 4.5 out of 5 stars38 ratings | 410.00 |
| Daughter of the Deep | Writtenby:RickRiordan | Narratedby:SoneelaNankani | 11 hrs and 16 mins | 05-10-21 | English | 4.5 out of 5 stars12 ratings | 615.00 |
| The Lightning Thief: Percy Jackson, Book 1 | Writtenby:RickRiordan | Narratedby:JesseBernstein | 10 hrs | 13-01-10 | English | 4.5 out of 5 stars181 ratings | 820.00 |
| The Hunger Games: Special Edition | Writtenby:SuzanneCollins | Narratedby:TatianaMaslany | 10 hrs and 35 mins | 30-10-18 | English | 5 out of 5 stars72 ratings | 656.00 |
| Quest for the Diamond Sword | Writtenby:WinterMorgan | Narratedby:LukeDaniels | 2 hrs and 23 mins | 25-11-14 | English | 5 out of 5 stars11 ratings | 233.00 |
| The Dark Prophecy | Writtenby:RickRiordan | Narratedby:RobbieDaymond | 12 hrs and 32 mins | 02-05-17 | English | 5 out of 5 stars50 ratings | 820.00 |
| Merlin Mission Collection | Writtenby:MaryPopeOsborne | Narratedby:MaryPopeOsborne | 10 hrs and 56 mins | 02-05-17 | English | 5 out of 5 stars5 ratings | 1,256.00 |
| The Tyrant's Tomb | Writtenby:RickRiordan | Narratedby:RobbieDaymond | 13 hrs and 22 mins | 24-09-19 | English | 5 out of 5 stars58 ratings | 820.00 |

# Steps to Create a Visualization

1. Define Goal and Audience
2. Understand and Prepare the Data
3. Choose the Right Visualization Type
4. Select the Right Tool
5. Design & Refine

datacamp

# The Data Visualization Cheat Sheet

Learn Data Visualization online at www.DataCamp.com

## How to use this cheat sheet

Use this cheat sheet for inspiration when making your next data visualizations. For more data visualization cheat sheets, check out our cheat sheets repository here.

## > Part-to-whole charts

### Pie chart

One of the most common ways to show part to whole data. It is also commonly used with percentages

USE CASES

1. Voting preference by age group
2. Market share of cloud providers

### Donut pie chart

The donut pie chart is a variant of the pie chart, the difference being it has a hole in the center for readability

USE CASES

1. Android OS market share
2. Monthly sales by channel

### Heat maps

Heatmaps are two-dimensional charts that use color shading to represent data trends.

USE CASES

1. Average monthly temperatures across the year
2. Departments with the highest amount of attrition over time

### Stacked column chart

Best to compare subcategories within categorical data. Can also be used to compare percentages

USE CASES

1. Quarterly sales per region
2. Total car sales by producer

### Treemap charts

2D rectangles whose size is proportional to the value being measured and can be used to display hierarchically structured data

USE CASES

1. Grocery sales count with categories
2. Stock price comparison by industry and company

## > Capture a trend

### Line chart

The most straightforward way to capture how a numeric variable is changing over time

USE CASES

1. Revenue in $ over time
2. Energy consumption in kWh over time
3. Google searches over time

### Multi-line chart

Captures multiple numeric variables over time. It can include multiple axes allowing comparison of different units and scale ranges

USE CASES

1. Apple vs Amazon stocks over time
2. Lebron vs Steph Curry searches over time
3. Bitcoin vs Ethereum price over time

### Area chart

Shows how a numeric value progresses by shading the area between line and the x-axis

USE CASES

1. Total sales over time
2. Active users over time

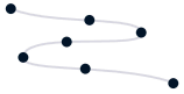
### Stacked area chart

Most commonly used variation of area charts, the best use is to track the breakdown of a numeric value by subgroups

USE CASES

1. Active users over time by segment
2. Total revenue over time by country

### Spline chart



Smoothened version of a line chart. It differs in that data points are connected with smoothed curves to account for missing values, as opposed to straight lines

USE CASES

1. Electricity consumption over time
2. CO2 emissions over time

## > Visualize a single value

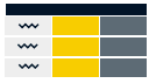
### Card


Cards are great for showing and tracking KPIs in dashboards or presentations

USE CASES

1. Revenue to date on a sales dashboard
2. Total sign-ups after a promotion

### Table chart



Best to be used on small datasets, it displays tabular data in a table

USE CASES

1. Account executive leaderboards
2. Registrations per webinar

### Gauge chart

This chart is often used in executive dashboard reports to show relevant KPIs

USE CASES

1. NPS scores
2. Revenue to target

## > Capture distributions

### Histogram

Shows the distribution of a variable. It converts numerical data into bins as columns. The x-axis shows the range, and the y-axis represents the frequency

USE CASES

1. Distribution of salaries in an organization
2. Distribution of height in one cohort

### Box plot

Shows the distribution of a variable using 5 key summary statistics—minimum, first quartile, median, third quartile, and maximum

USE CASES

1. Gas efficiency of vehicles
2. Time spent reading across readers

### Violin plot

A variation of the box plot. It also shows the full distribution of the data alongside summary statistics

USE CASES

1. Time spent in restaurants across age groups
2. Length of pill effects by dose

### Density plot

Visualizes a distribution by using smoothing to allow smoother distributions and better capture the distribution shape of the data

USE CASES

1. Distribution of price of hotel listings
2. Comparing NPS scores by customer segment

## > Visualize relationships

### Bar chart

One of the easiest charts to read which helps in quick comparison of categorical data. One axis contains categories and the other axis represents values

### Column chart

Also known as a vertical bar chart, where the categories are placed on the x-axis. These are preferred over bar charts for short labels, date ranges, or negatives in values

### Scatter plot

Most commonly used chart when observing the relationship between two variables. It is especially useful for quickly surfacing potential correlations

### Connected scatterplot

A hybrid between a scatter plot and a line plot, the scatter dots are connected with a line

### Bubble chart

Often used to visualize data points with 3 dimensions, namely visualized on the x-axis, y-axis, and with the size of the bubble. It tries to show relations between data points

### Word cloud chart


A convenient visualization for visualizing the most prevalent words that appear in a text

## > Visualize a flow

### Sankey chart

Useful for representing flows in systems. This flow can be any measurable quantity

### Chord chart

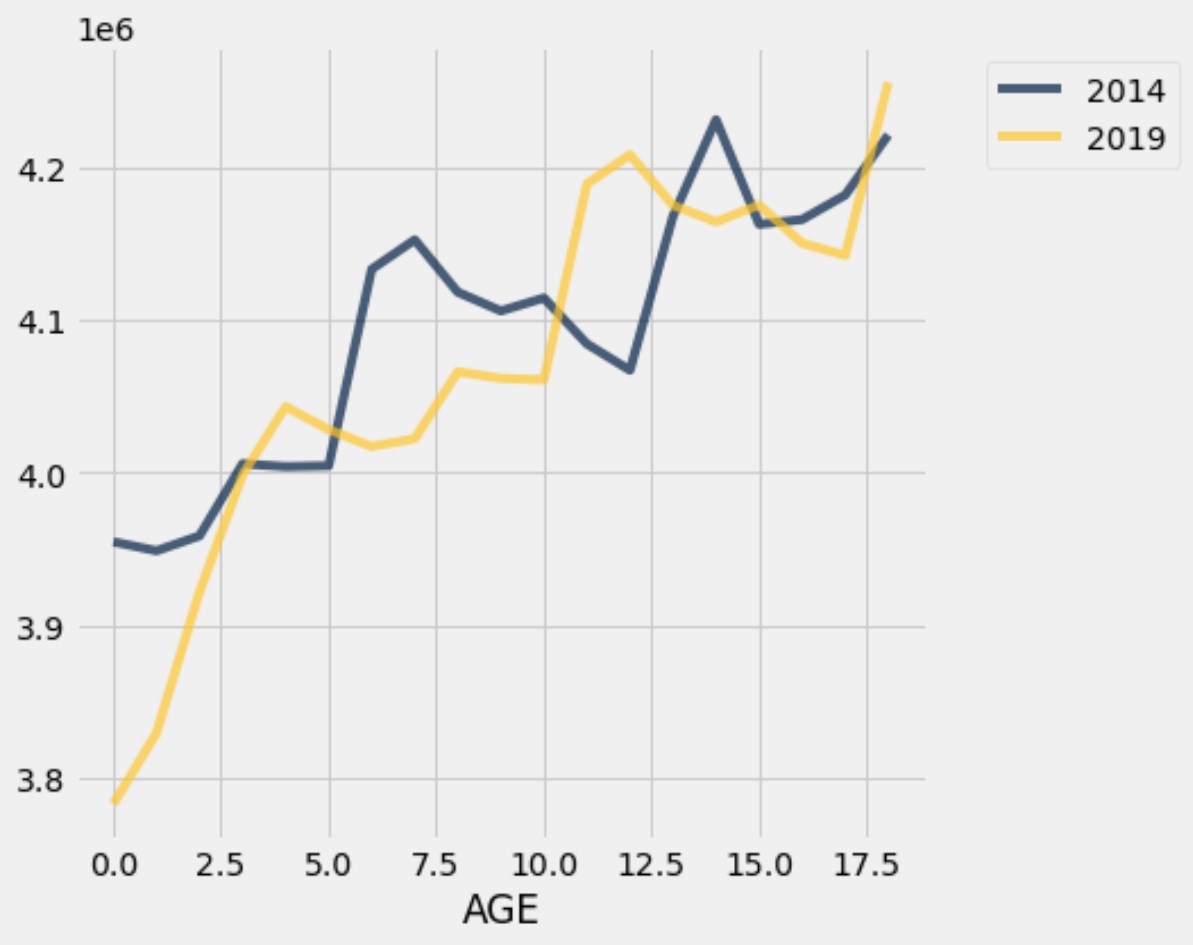
Useful for presenting weighted relationships or flows between nodes. Especially useful for highlighting the dominant or important flows

### Network chart

Similar to a graph, it consists of nodes and interconnected edges. It illustrates how different items have relationships with each other

Learn Data Skills Online at
www.DataCamp.com

# Line Graphs



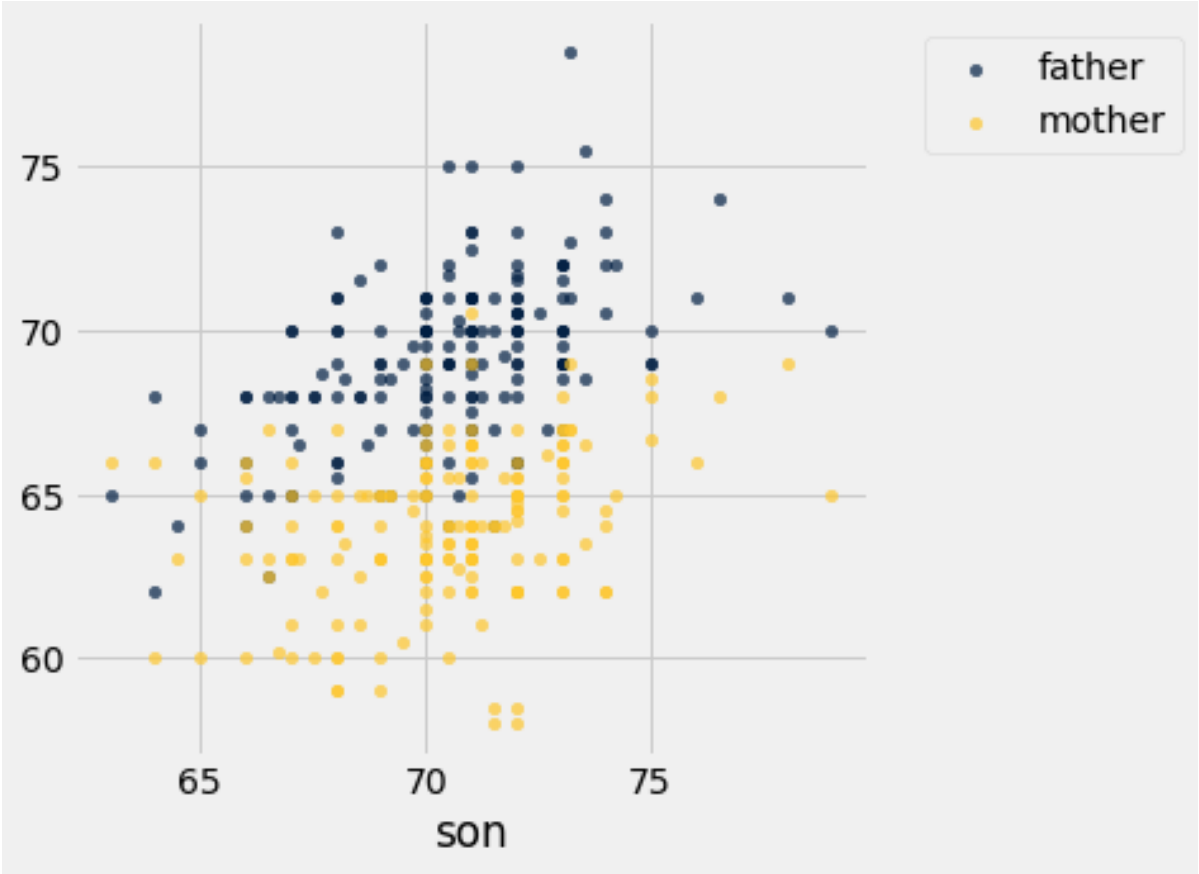

- Uses:
  - Trends
  - Especially for looking at chronological trends!
- Things to consider:
  - Multiple lines
  - Time intervals
  - Truncating axes
- Examples:
  - Census Data
  - Monthly sales revenue

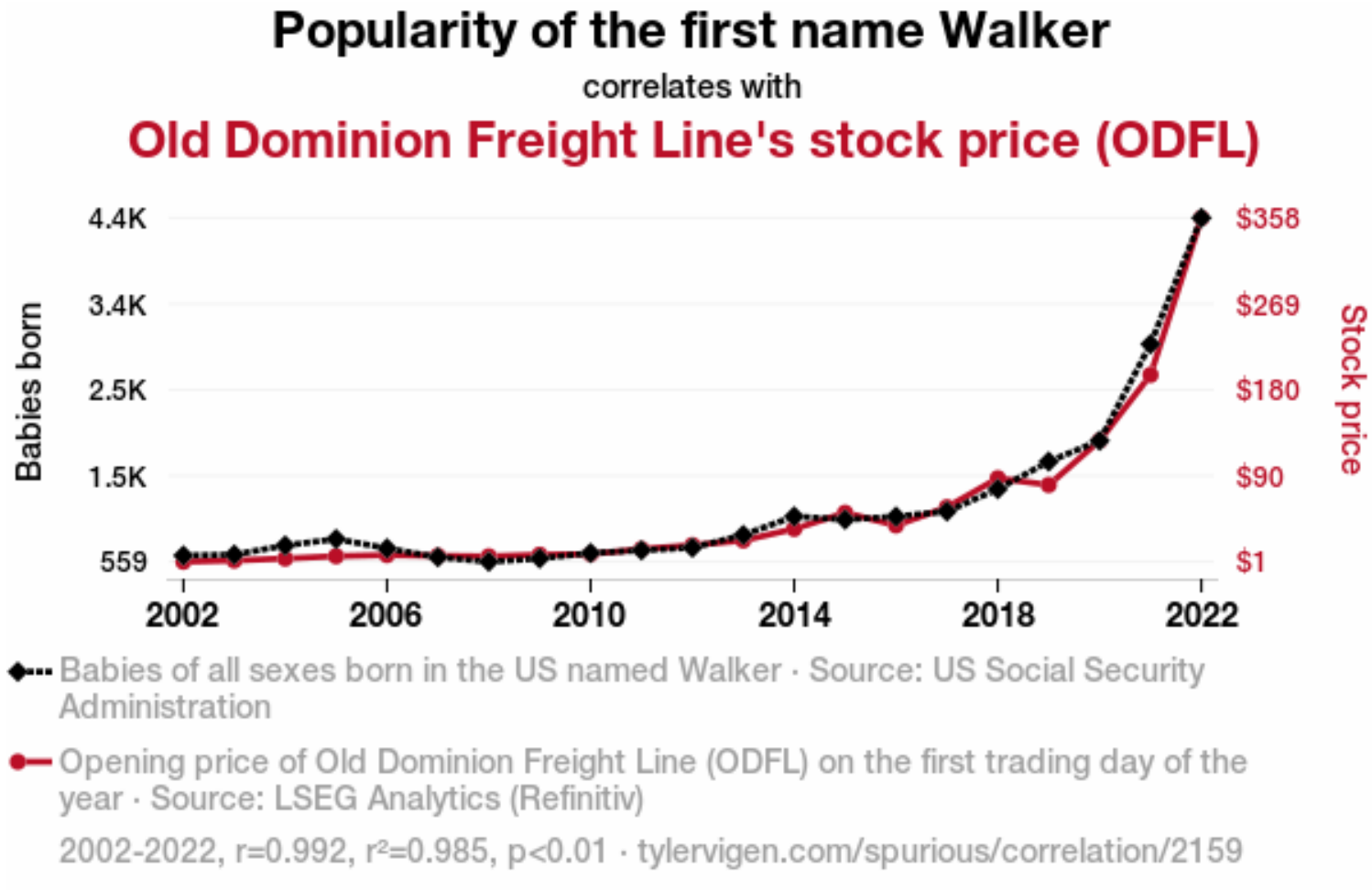
Let's look at indicators of world environmental health example.

DEMO

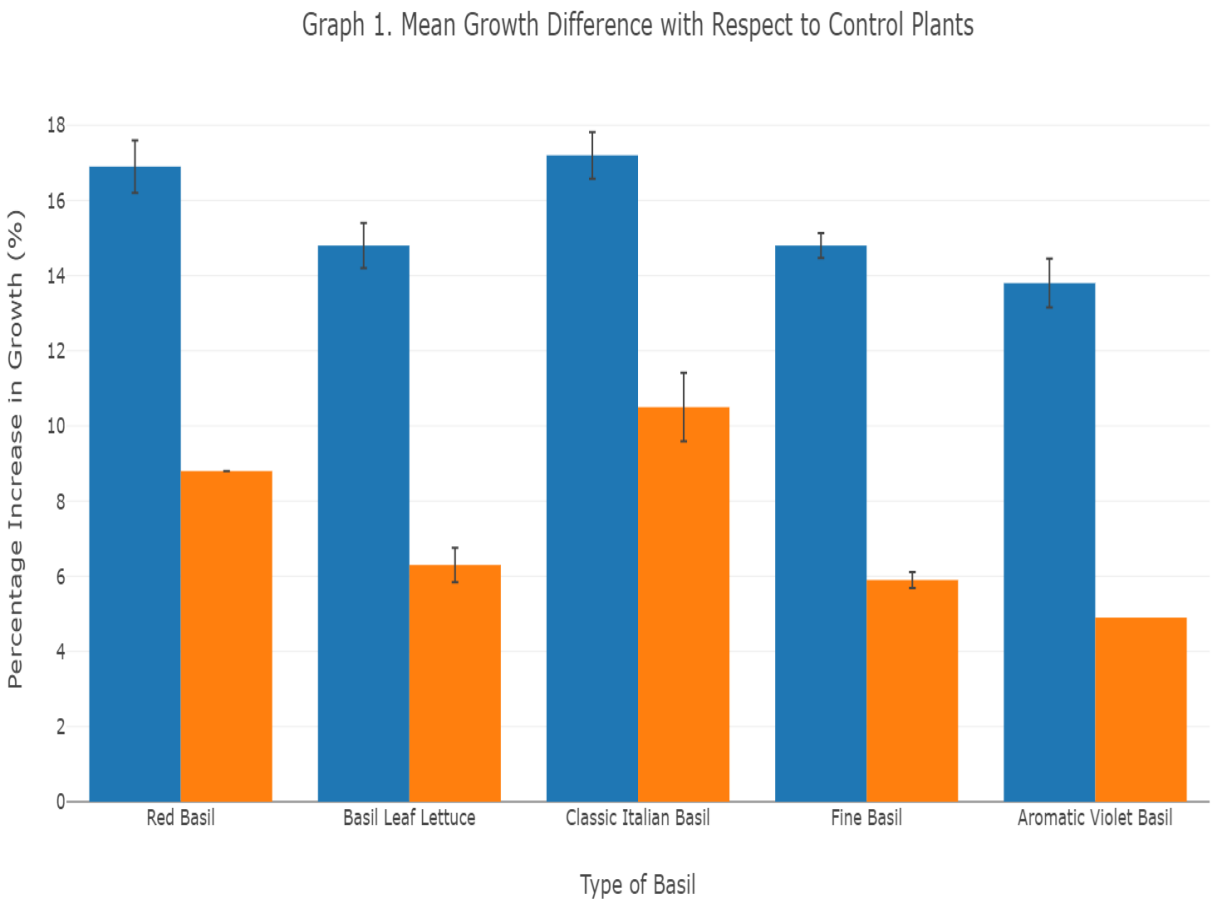
# Scatter Plots

- Uses:
  - Relationship between two numerical variables
- Things to Consider:
  - Trendline
  - Overplotting – consider size and transparency
  - Correlation does not equal causation
- Examples:
  - Parent Height to Child Height
  - Budget versus revenue




Popularity of the first name Walker
correlates with
Old Dominion Freight Line's stock price (ODFL)
Babies born
4.4K
3.4K
2.5K
1.5K
559
$358
$269
$180
$90
$1
Stock price
2002
2006
2010
2014
2018
2022
Babies of all sexes born in the US named Walker · Source: US Social Security Administration
Opening price of Old Dominion Freight Line (ODFL) on the first trading day of the year · Source: LSEG Analytics (Refinitiv)
2002-2022, r=0.992, r²=0.985, p<0.01 · tylervigen.com/spurious/correlation/2159

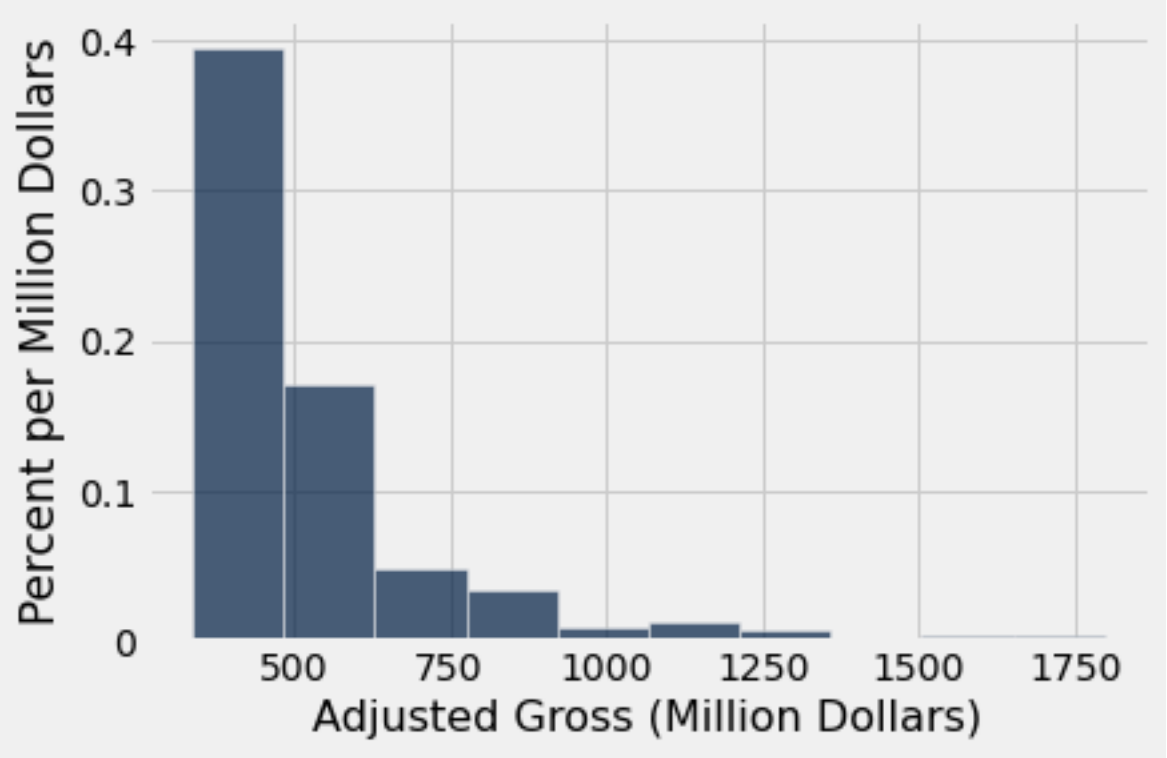
# Bar Charts



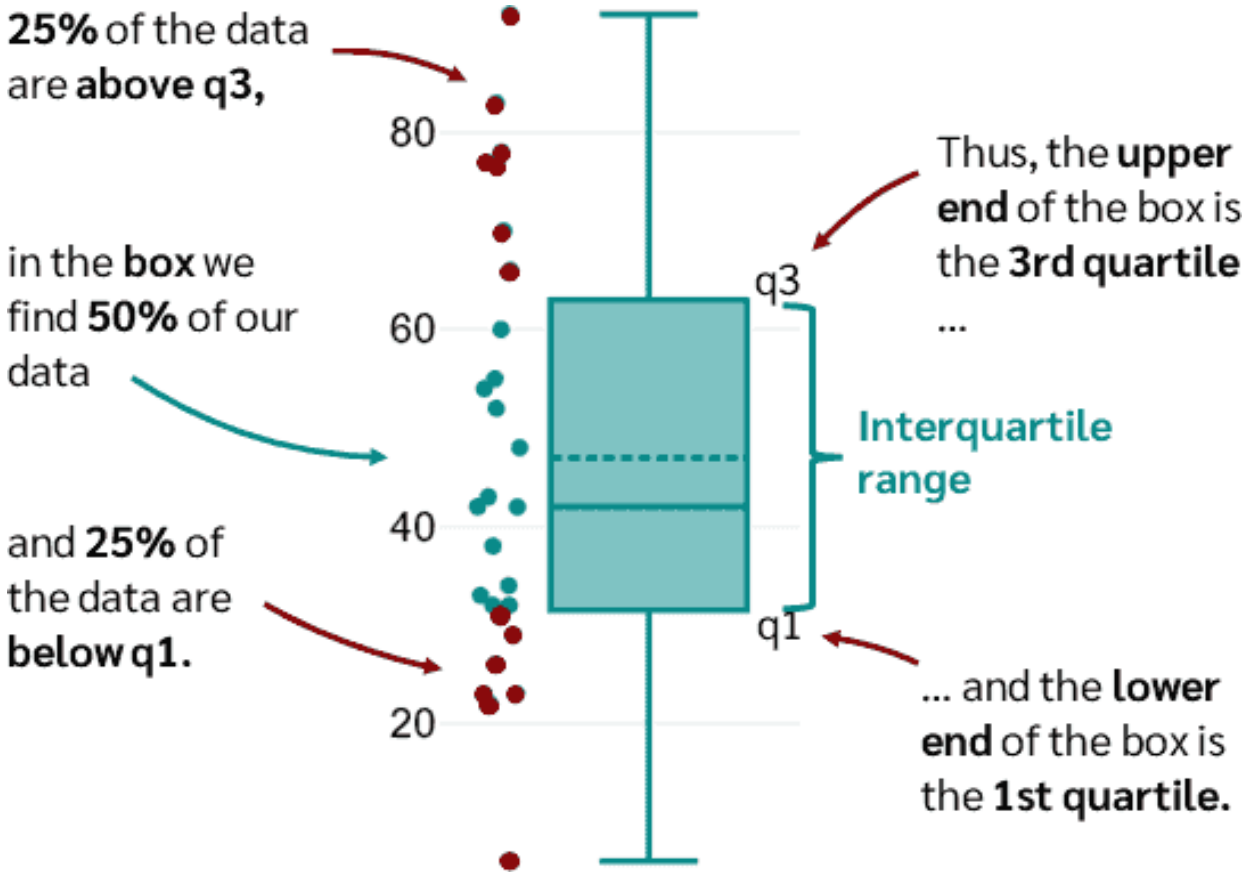

- Uses:
  - Comparing discrete categories (categorical data)
  - Ranking or magnitudes
- Things to Consider:
  - Bars should start at zero – truncate carefully!
  - Horizontal or vertical
- Examples:
  - Student test scores by subject
  - Mean growth by plant type

# Histograms

- Uses:
  - Understanding how values are distributed
  - Identifying skewness, gaps, unusual patterns
- Things to Consider:
  - Bin sizes are very important!
- Examples:
  - Housing prices in an area
  - Revenue of business in NY state



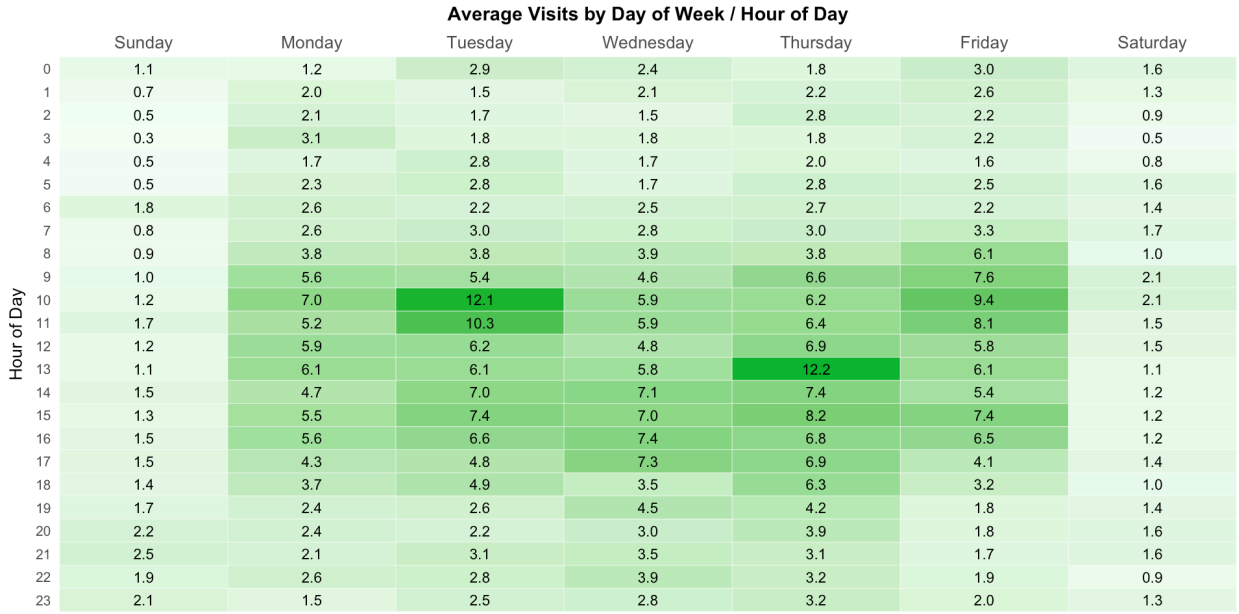
# Box Plot




- Uses:
  - Comparing distributions across groups
  - Identifying outliers and the spread of the data
- Things to Consider:
  - Moderate sample sizes – too few wouldn't be informative
  - Outliers
- Examples:
  - Comparing test scores across sections of a class

# Heat Maps

- Uses:
  - Showing relationships between multiple variables
  - Matrices
- Things to consider:
  - Color theme very important here!
  - Correlation does not equal causation
- Examples:
  - Website visits per day



**Average Visits by Day of Week / Hour of Day**

| Hour of Day | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
|---|---|---|---|---|---|---|---|
| 0 | 1.1 | 1.2 | 2.9 | 2.4 | 1.8 | 3.0 | 1.6 |
| 1 | 0.7 | 2.0 | 1.5 | 2.1 | 2.2 | 2.6 | 1.3 |
| 2 | 0.5 | 2.1 | 1.7 | 1.5 | 2.8 | 2.2 | 0.9 |
| 3 | 0.3 | 3.1 | 1.8 | 1.8 | 1.8 | 2.2 | 0.5 |
| 4 | 0.5 | 1.7 | 2.8 | 1.7 | 2.0 | 1.6 | 0.8 |
| 5 | 0.5 | 2.3 | 2.8 | 1.7 | 2.8 | 2.5 | 1.6 |
| 6 | 1.8 | 2.6 | 2.2 | 2.5 | 2.7 | 2.2 | 1.4 |
| 7 | 0.8 | 2.6 | 3.0 | 2.8 | 3.0 | 3.3 | 1.7 |
| 8 | 0.9 | 3.8 | 3.8 | 3.9 | 3.8 | 6.1 | 1.0 |
| 9 | 1.0 | 5.6 | 5.4 | 4.6 | 6.6 | 7.6 | 2.1 |
| 10 | 1.2 | 7.0 | 12.1 | 5.9 | 6.2 | 9.4 | 2.1 |
| 11 | 1.7 | 5.2 | 10.3 | 5.9 | 6.4 | 8.1 | 1.5 |
| 12 | 1.2 | 5.9 | 6.2 | 4.8 | 6.9 | 5.8 | 1.5 |
| 13 | 1.1 | 6.1 | 6.1 | 5.8 | 12.2 | 6.1 | 1.1 |
| 14 | 1.5 | 4.7 | 7.0 | 7.1 | 7.4 | 5.4 | 1.2 |
| 15 | 1.3 | 5.5 | 7.4 | 7.0 | 8.2 | 7.4 | 1.2 |
| 16 | 1.5 | 5.6 | 6.6 | 7.4 | 6.8 | 6.5 | 1.2 |
| 17 | 1.5 | 4.3 | 4.8 | 7.3 | 6.9 | 4.1 | 1.4 |
| 18 | 1.4 | 3.7 | 4.9 | 3.5 | 6.3 | 3.2 | 1.0 |
| 19 | 1.7 | 2.4 | 2.6 | 4.5 | 4.2 | 1.8 | 1.4 |
| 20 | 2.2 | 2.4 | 2.2 | 3.0 | 3.9 | 1.8 | 1.6 |
| 21 | 2.5 | 2.1 | 3.1 | 3.5 | 3.1 | 1.7 | 1.6 |
| 22 | 1.9 | 2.6 | 2.8 | 3.9 | 3.2 | 1.9 | 0.9 |
| 23 | 2.1 | 1.5 | 2.5 | 2.8 | 3.2 | 2.0 | 1.3 |

# Comparing Bar Charts and Histograms

| | **BAR CHART** | **HISTOGRAM** |
|---|---|---|
| **Type of Data** | Categorical Variables | Numerical variables |
| **Axis** | 1 categorical & 1 numerical | Both numerical axes |
| **Width of Bars** | Arbitrary but Equal | To scale with no gaps |
| **Height of Bar** | Proportional to percent of individuals | Density of bar |
| **Area of Bar** | N/A | Proportional to percent of individuals |

# How to pick the visualization type?

| Title | Author | Description |
|---|---|---|
| **Chart Chooser** | Ann K Emery | (interactive) Small/Focused collection with commentary and resources. See her design process. |
| **The Data Visualization Catalogue** | Severino Ribecca | (interactive) Large collection of all visualizations, plus description, details "anatomy", similar charts, and relevant software/tools |
| **Data Viz Project** | ferdio | (interactive) Large collection of all visualizations with filters by data structure |
| **from Data to Viz** | Yan Holtz & Conor Healy | (interactive) Medium collection of graph types with filters by data structure or purpose, and links to R and Python galleries. |
| **Chartopedia** | AnyChart JS Library | (interactive) Includes example with complete html, css, data, and javscript |
| **Chart Chooser** | Tableau | (interactive) Includes downloadable Tableau workbook and instructions |
| **Visual Vocabulary** | Financial Times | Nice poster with a well-organized large collection. Web page has resources and translations.The poster that inspired it cannot be seen well without purchase. See examples in Tableau |
| **Chart Chooser** | Kagen Quiballo | A small collection of the most popular charts organized by name and goal, with links to videos about creating in Tableau and Excel. |
| **Chooser for Trends** | MetaPraxix | Focused collection with typical challenges for trends. |
| **Infographic Taxonomy** | Frederik Ruys, Vizualism | Matrix for infographics that express a combination of the questions Who, When, Where, What, How, and How Much. |
| **Qualitative Chart Chooser** | Stephanie Evergreen | Matrix of non-typical chart types and purposes (i.e., no bar charts) |
| **SWD Chart Guide** | Storytelling with Data | Standard chart types with descriptions and tips for creating. |
| **Visualization Universe** | Adioma | Medium collection showing popularity with links to related terms and the Wikipedia page. |

# Steps to Create a Visualization

1. Define Goal and Audience
2. Understand and Prepare the Data
3. Choose the Right Visualization Type
4. Select the Right Tool
5. Design & Refine

# Software

- Non-coding Tools
  - Excel and Google Sheets
- Coding Tools
  - Python: Matplotlib, Plotly, Seaborn
  - R: ggplot2
- Online Tools
  - Tableau (https://public.tableau.com/app/discover) Free!
  - Power BI
  - D3.js (https://d3js.org/)

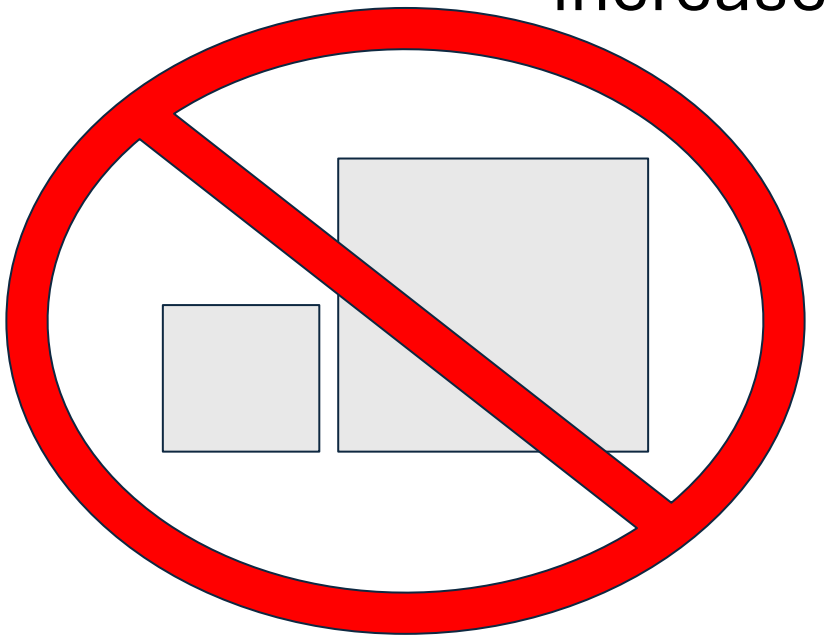
# Steps to Create a Visualization

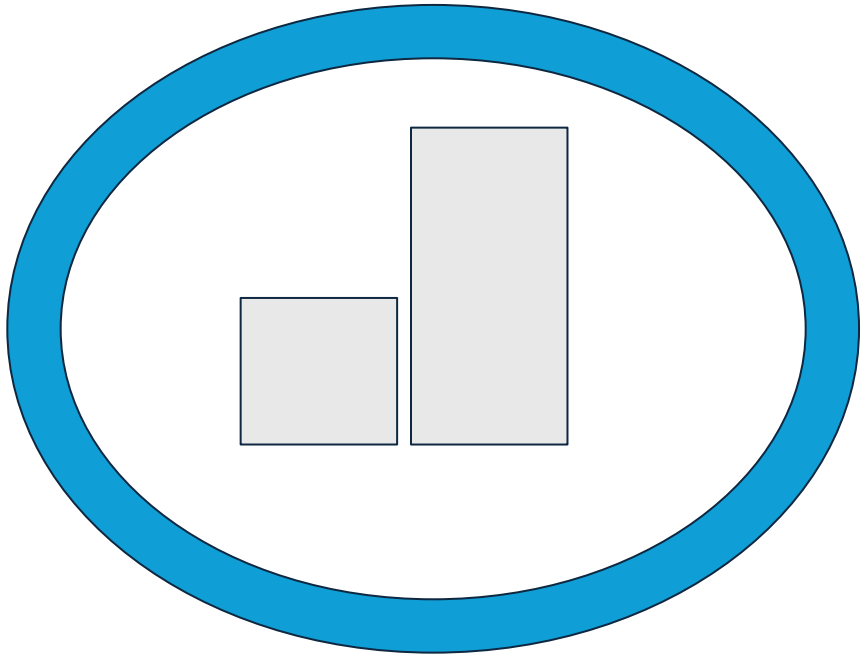
1. Define Goal and Audience
2. Understand and Prepare the Data
3. Choose the Right Visualization Type
4. Select the Right Tool
5. Design & Refine

# Area Principle

Keep in mind!  Areas should be proportional to the values that they represent.

If we want to show an increase between two items, we should NOT increase in both dimensions.





# Principles to keep in mind

- Color Theory and Accessibility
  - Use contrasting colors
  - Keep accessibility in mind
  - Lots of tools out there: https://coolinfographics.com/color-pickers
- Tufte's Principles
  - Graphical Integrity
  - Chart Junk
  - Documentation
  - Context