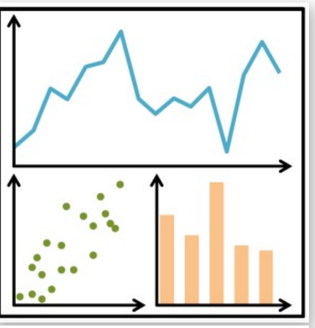
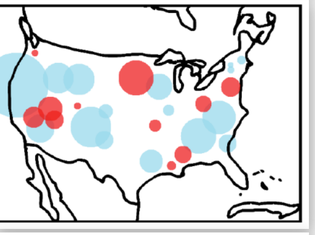
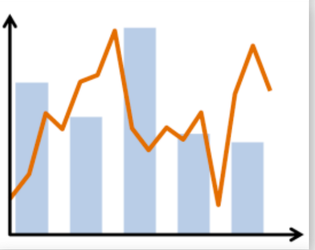
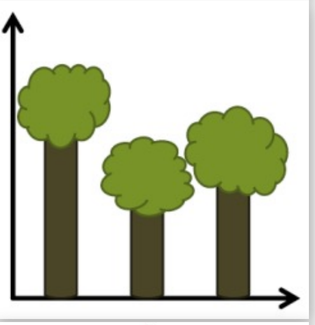
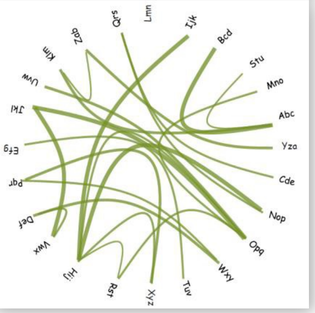






Lecture 1: Introduction, Course Overview, Git

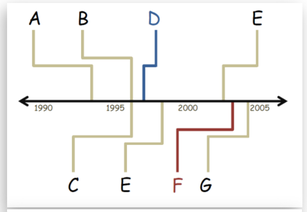
DS 4200

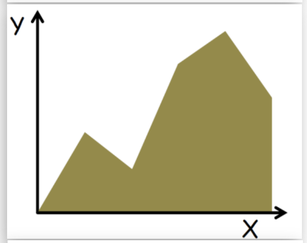
SPRING 2023

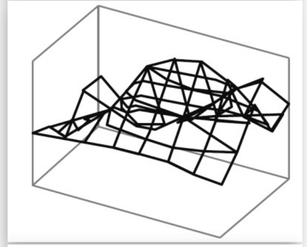
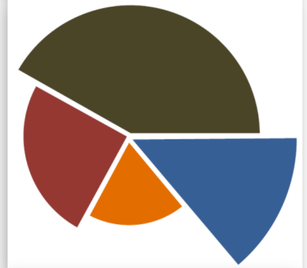
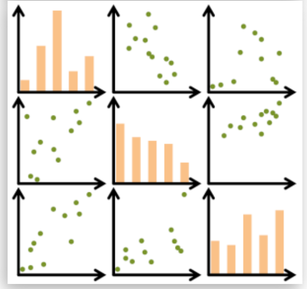
Prof. Ab Mosca (they/them)

NORTHEASTERN UNIVERSITY

Slides and inspiration from Cody Dunne, Michelle Borkin, Dylan Cashman, Krzysztof Gajos, Hanspeter Pfister, Miriah Meyer, Jonathan Schwabish, and David Sprague





Today

- Course Overview
- What is Visualization? Why Visualization?
- Intro to Git

Teaching Staff Introductions

Professor:

Ab Mosca (they/them*)
PhD, Computer Science

*I use they/them pronouns. Yes, it is inappropriate to use any other pronouns for me.
If you would like resources on pronoun usage and/or a place to practice, try starting here: http://showyourlovetoday.com/wp-content/uploads/2018/09/Pronoun-Practice.pdf

Office Hours:

Tue. 1:00pm – 3:00pm, and
Wed. 2:30pm – 5:00pm

Book a 15 min. time slot using Calendly*: https://calendly.com/aemosca/officehours.

If you cannot make it to my scheduled hours, please email to set up another time: a.mosca@northeastern.edu

*You must book at least 1 hour in advance.

Teaching Staff Introductions

TA's:

Abigail Sodgergren
(she/her)

Ashraf Bade
(he/him)

Luca Sharbani
(they/he)

Khushi Morparia
(she/her)

Office Hours:

See Course Website for most up to date information.

Teaching Staff Introductions

Extra Office Hours with Kate Kryder at Snell Library

Kate can help with:

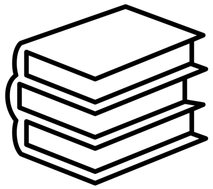
- Ideas for visualizing your data or topic
- Recommendations for data visualization tools to use
- Assistance using data visualization tools (ex. R, Python, D3, Tableau)
- Suggestions for improving data visualization design
- Poster design
- Presentation design

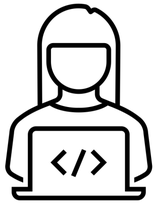
Book a time with Kate here: https://subjectguides.lib.neu.edu/dataviz/consults

Course Overview

- **Overarching goal:** Learn to create effective information representations
- How will we accomplish this?

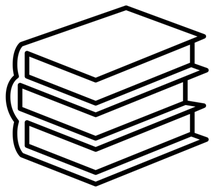
Theory + Application

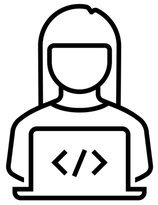


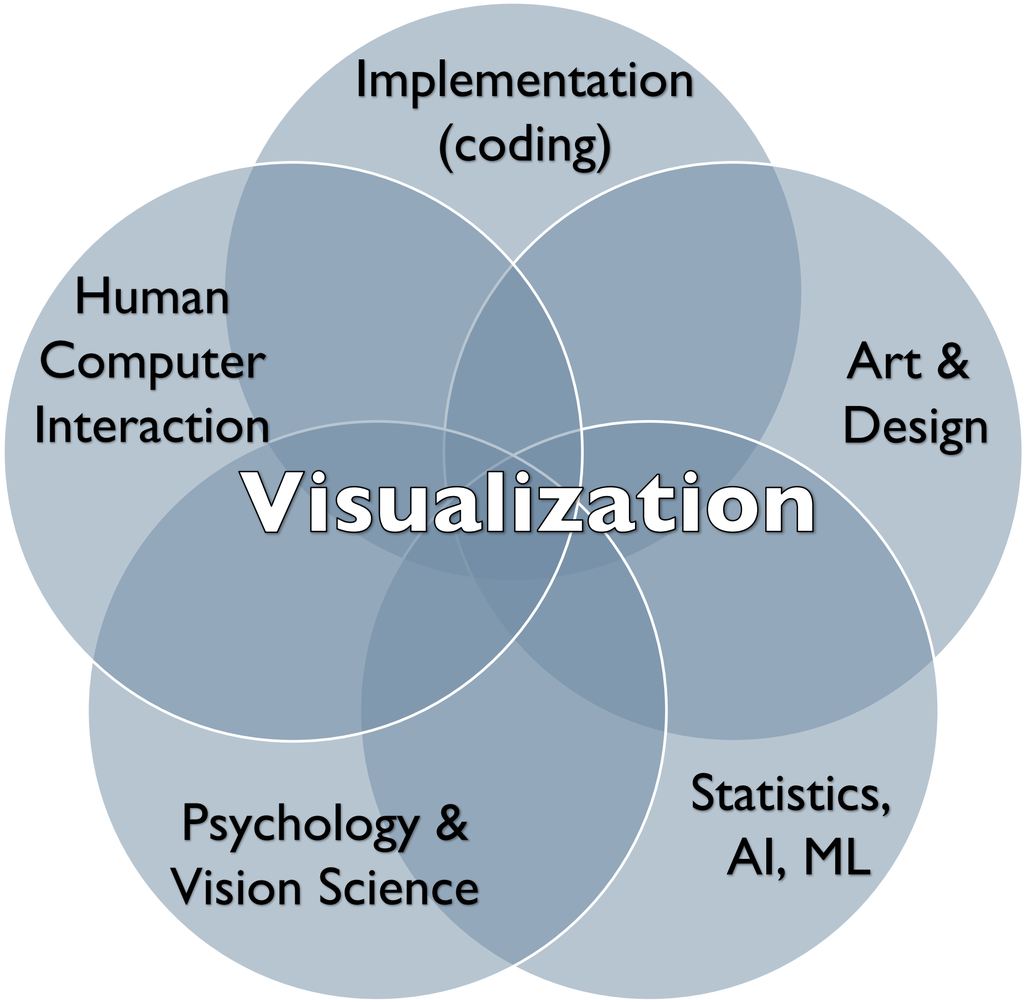


Course Overview

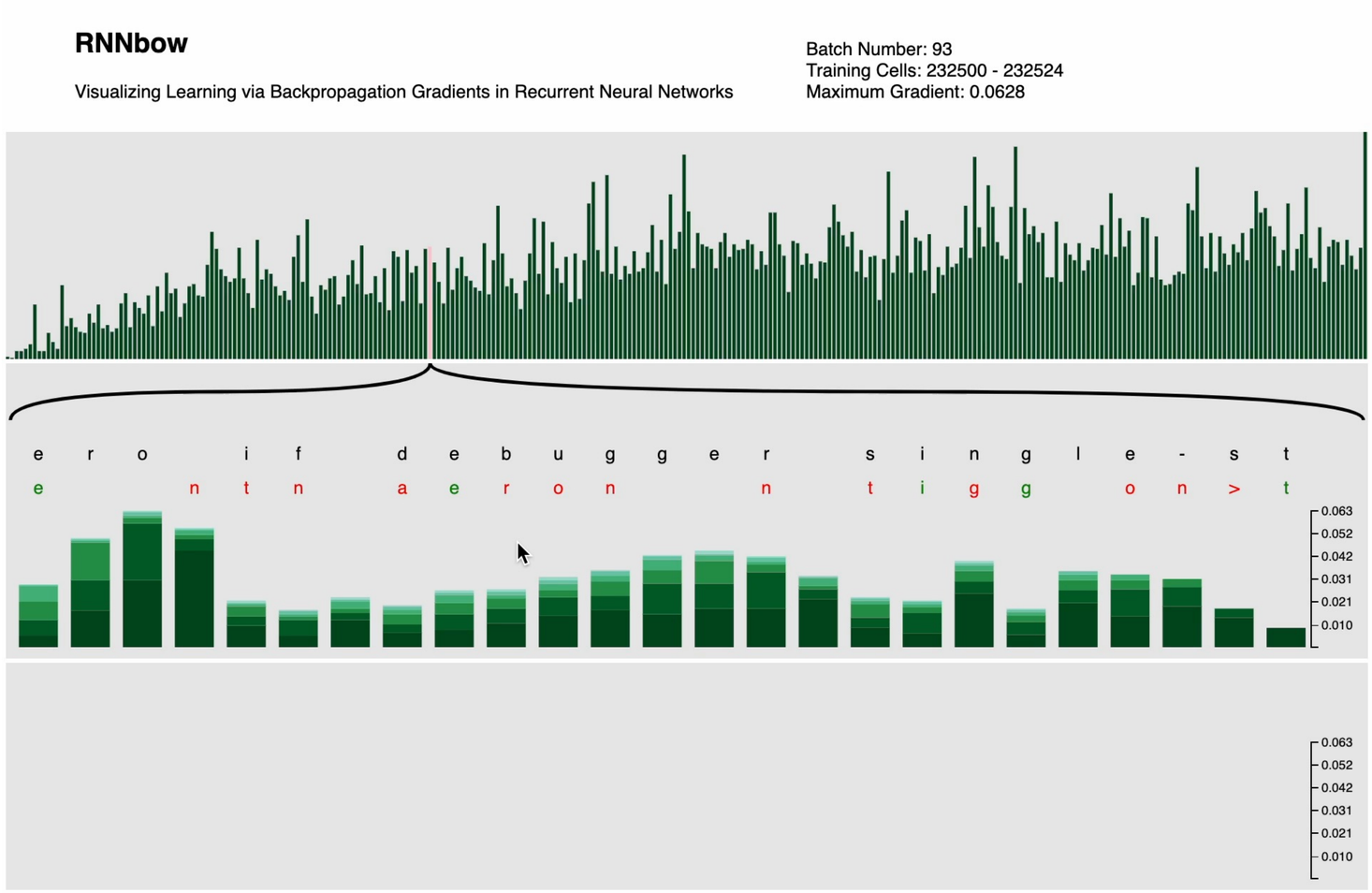
Theory + Application



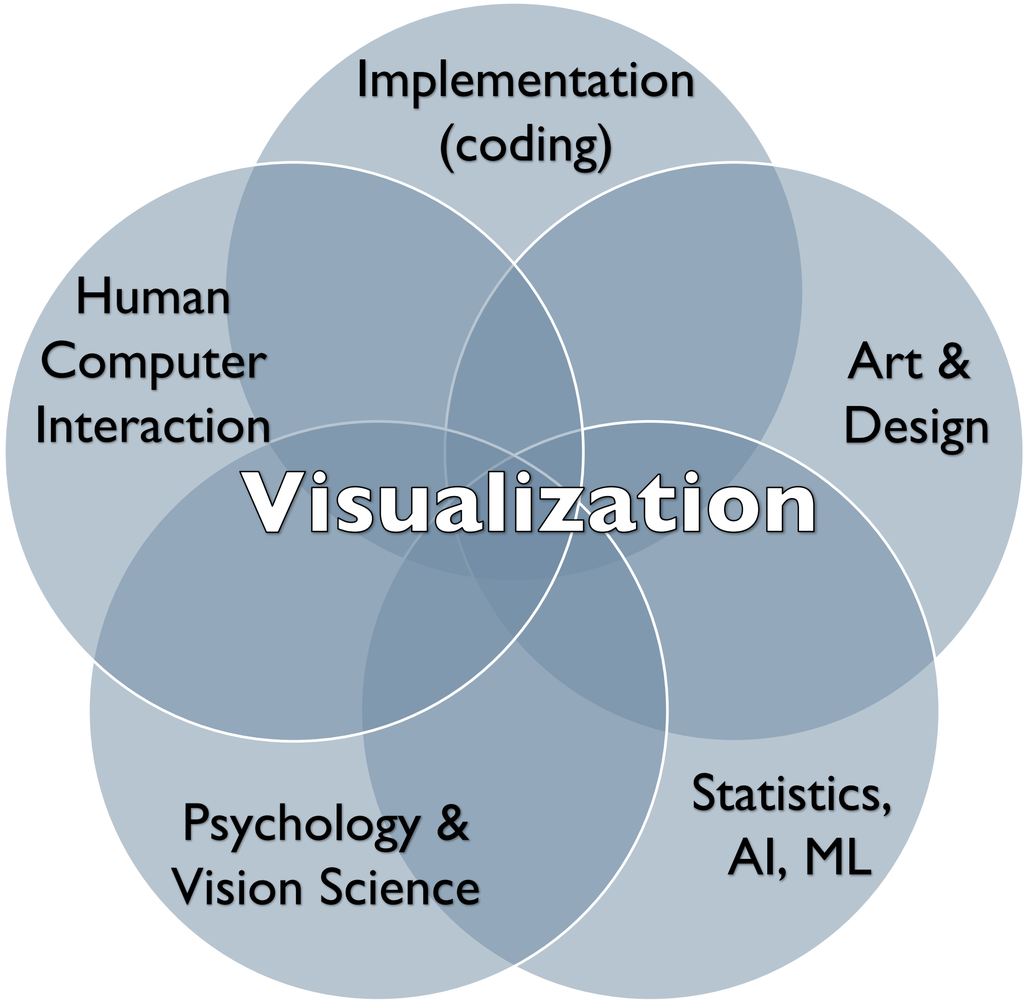




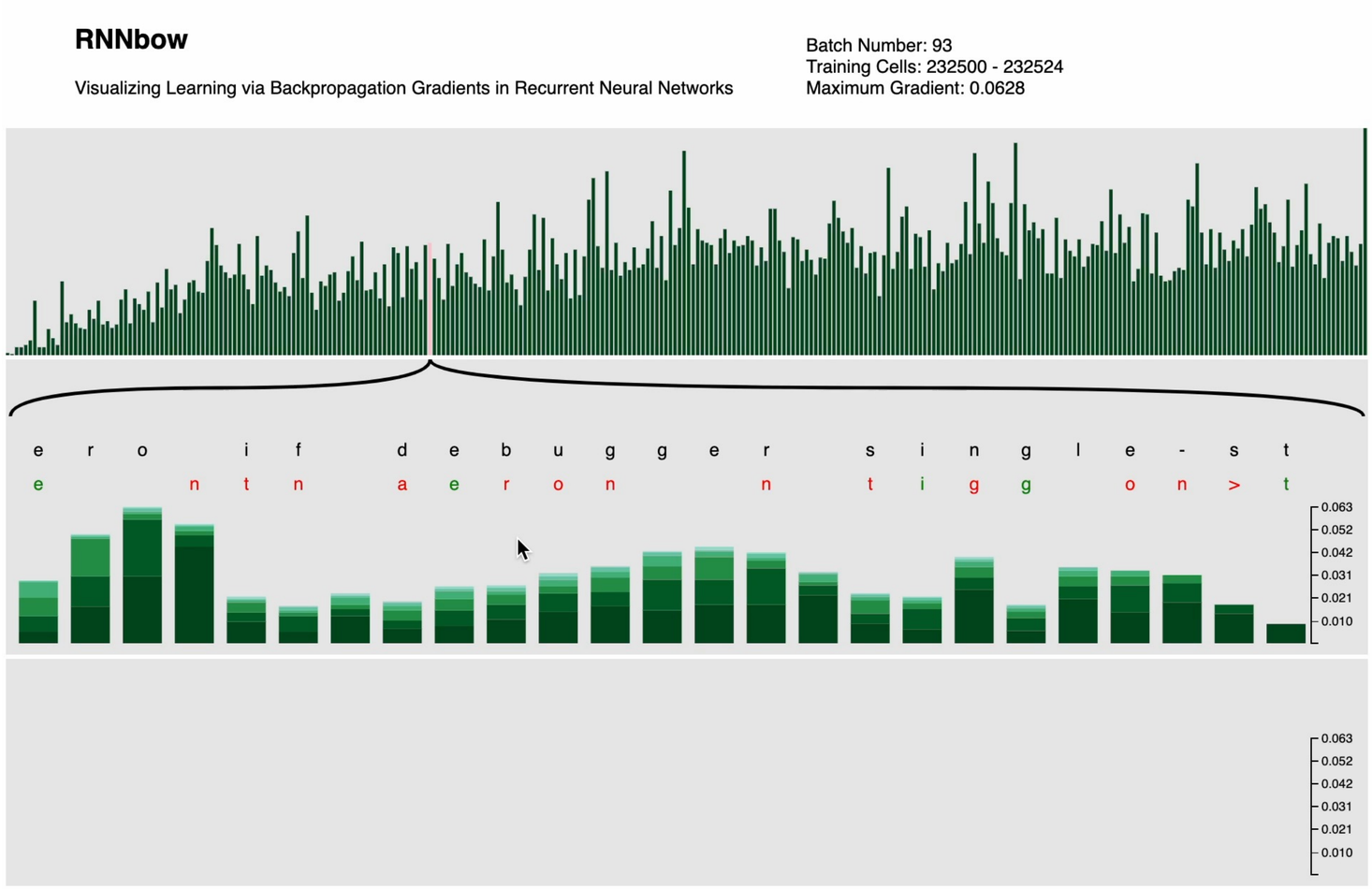
Course Overview



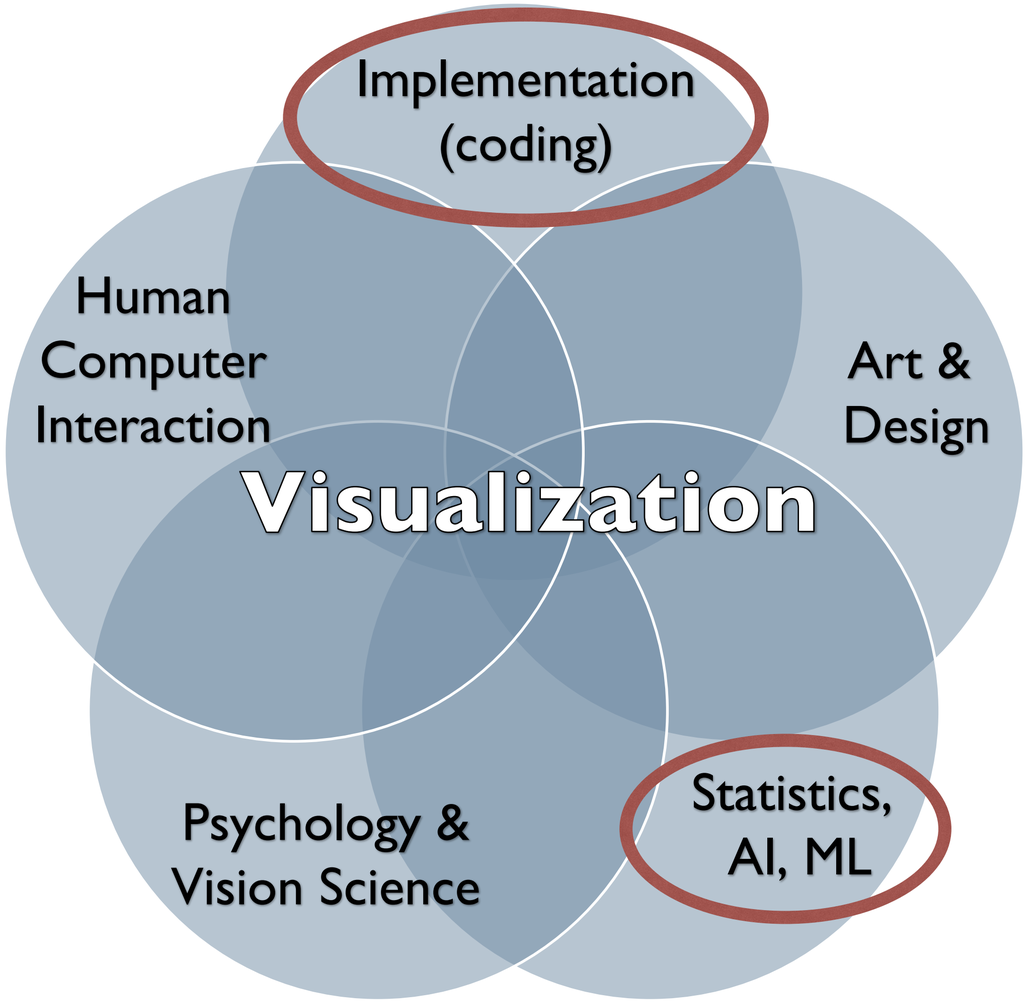
https://vimeo.com/375164049



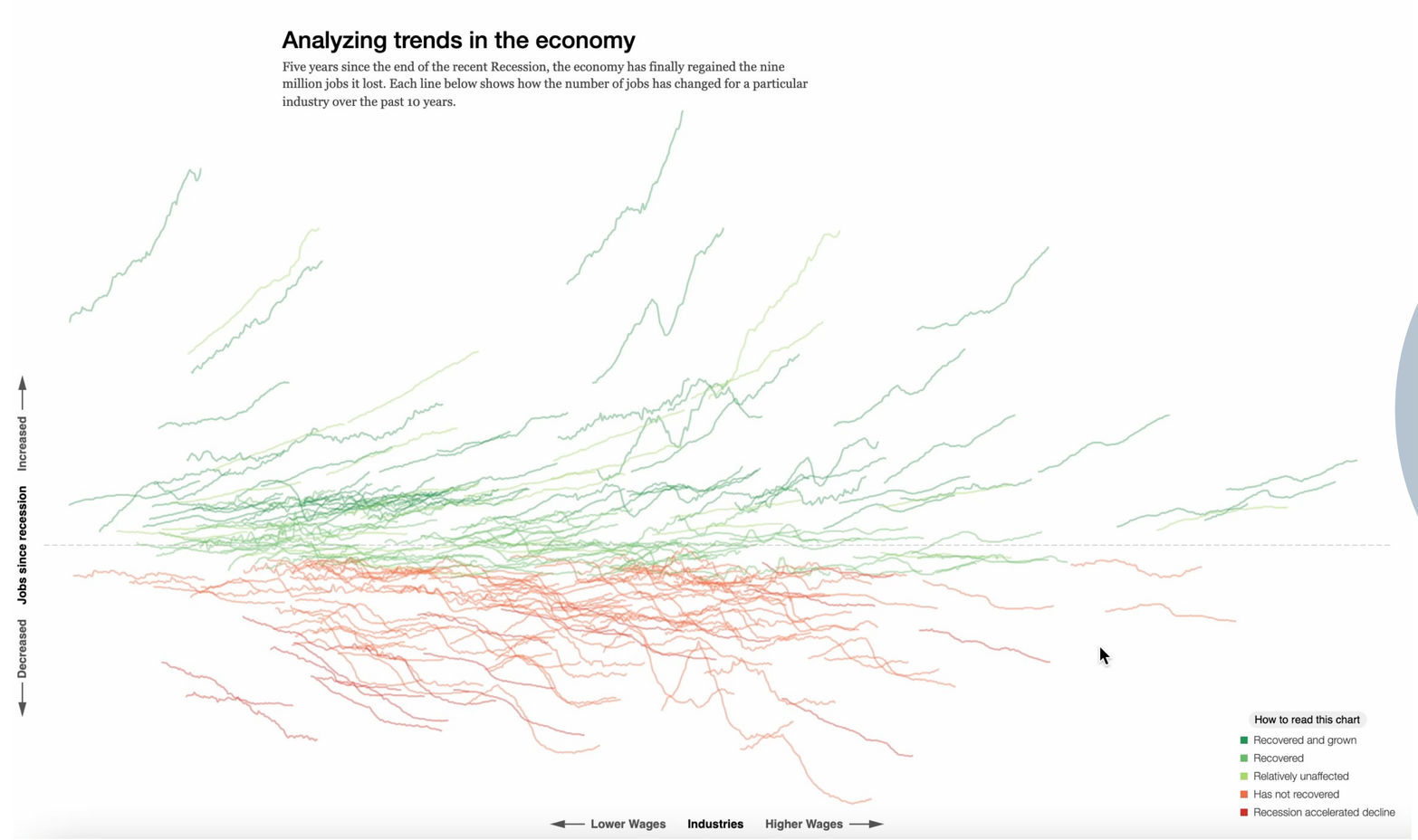
Course Overview



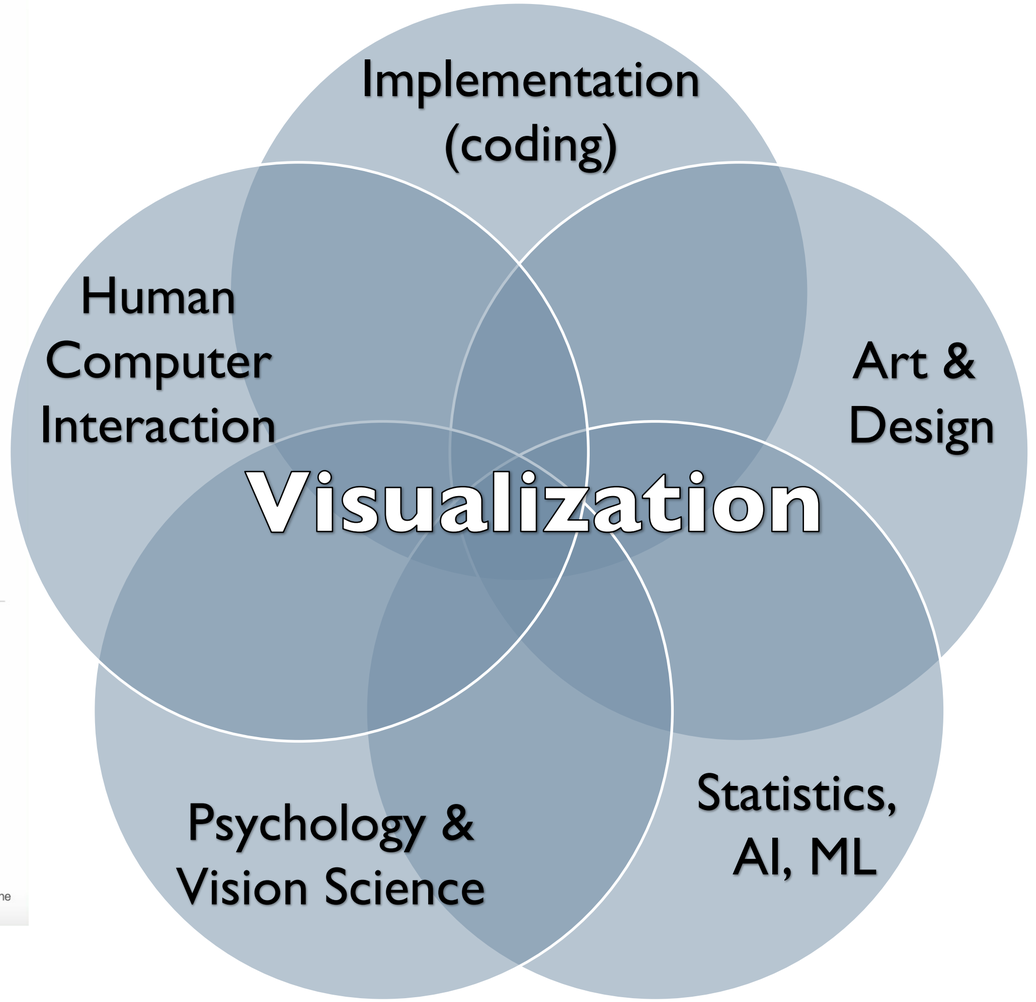
https://vimeo.com/375164049



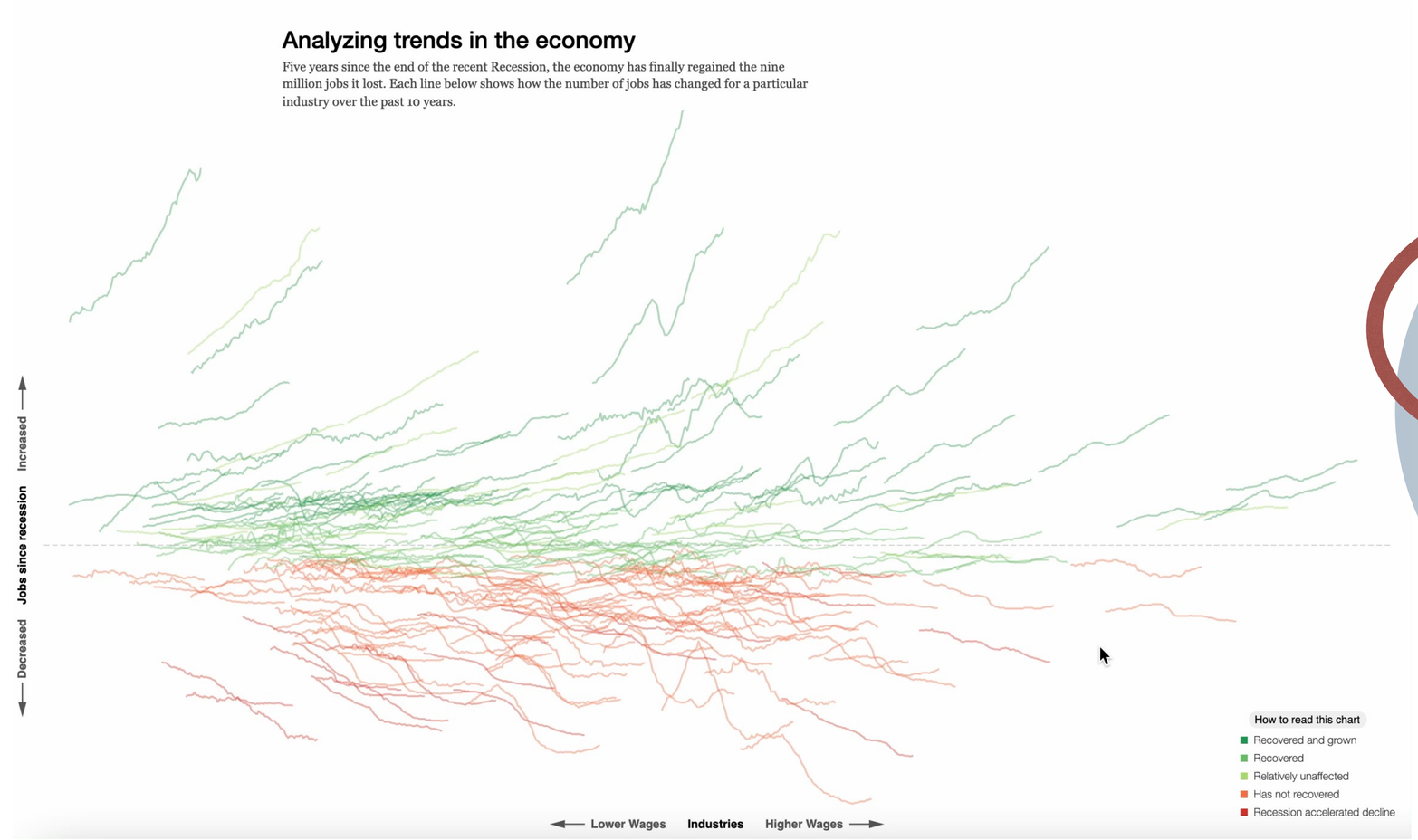
Course Overview



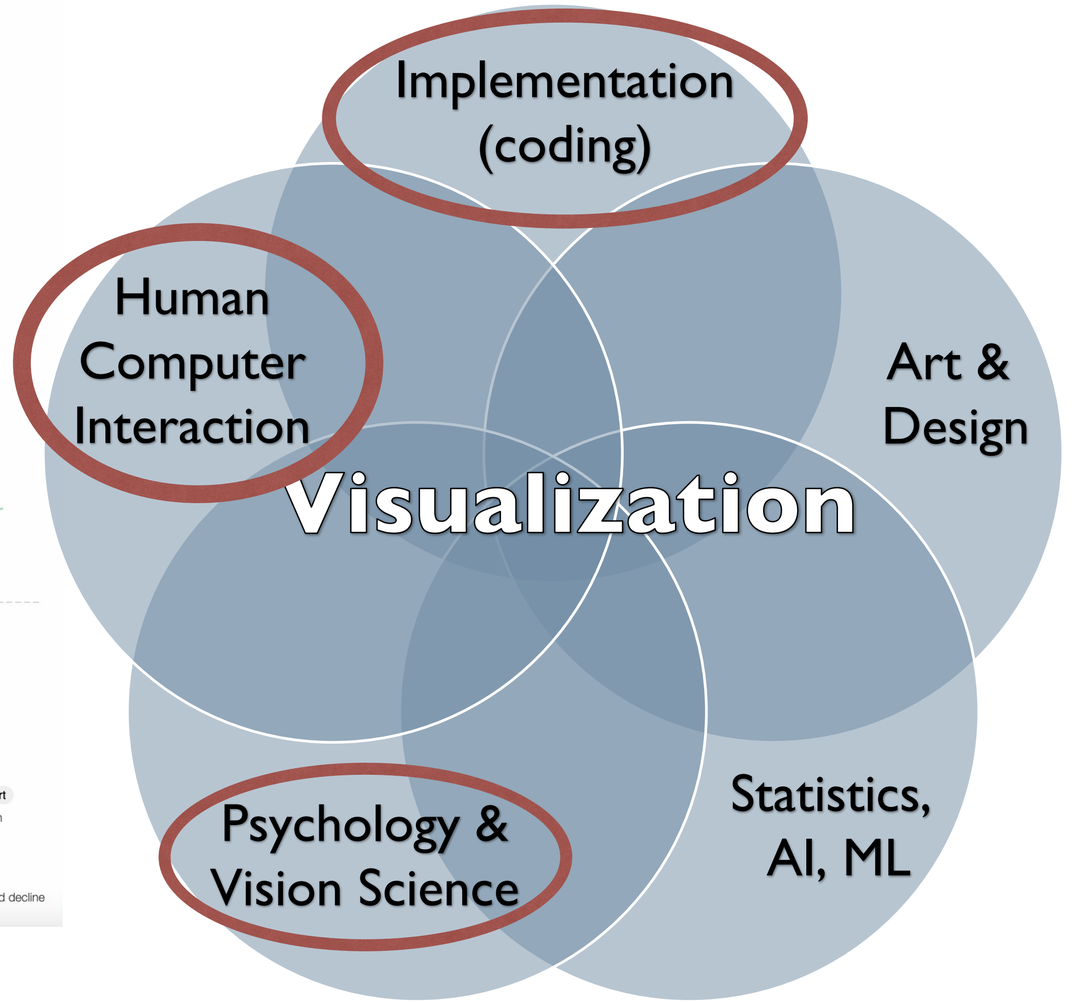
https://medium.com/bucknell-hci/hindsight-encouraging-exploration-and-engagement-in-data-visualization-7b0d9d421fc5#.n0l56dson



Course Overview

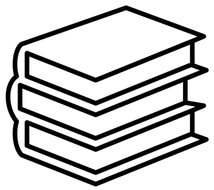


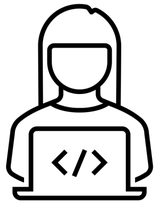
https://medium.com/bucknell-hci/hindsight-encouraging-exploration-and-engagement-in-data-visualization-7b0d9d421fc5#.n0l56dson

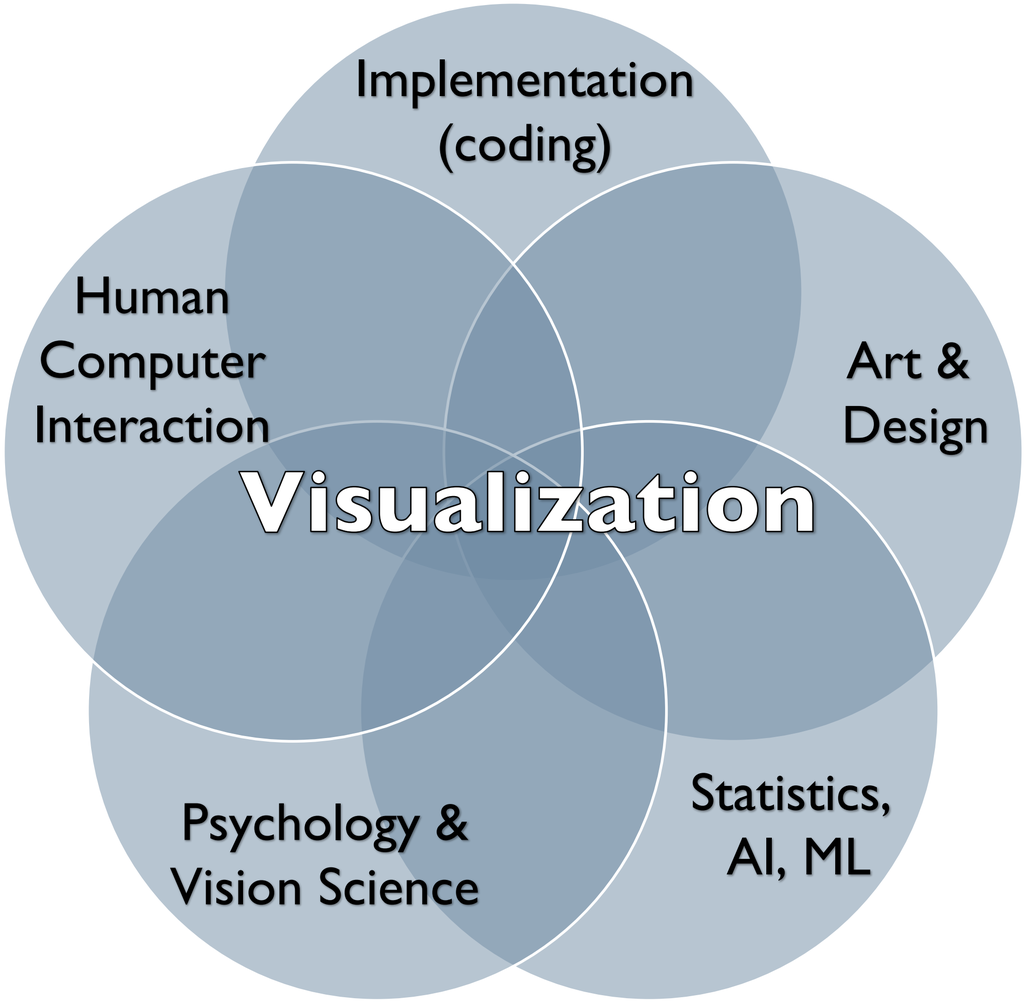


Course Overview

Theory + Application







Course Structure

Classroom Technologies

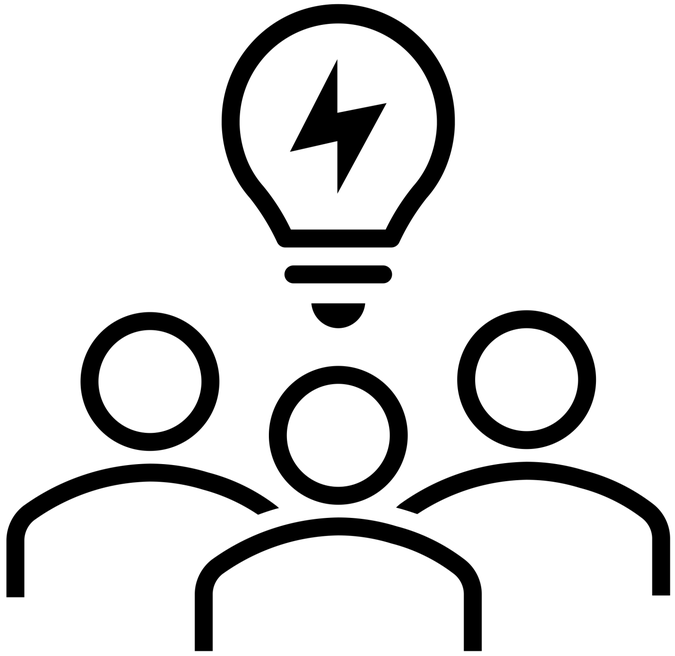
- Course Website - [httpshttps://amosca01.github.io/DS4200-S23/](https://amosca01.github.io/DS4200-S23/)
- Piazza - https://piazza.com/class/lcge9ggpbdf7eu
- Gradescope - https://www.gradescope.com/courses/480889 (entry code BBRJRX)
- GitHub Classroom – links provided in assignments

Class Structure

- In class
  - Lecture + graded in-class activity (ic)
- Outside of class
  - Graded homework (hw) + project milestone (pm) assignments

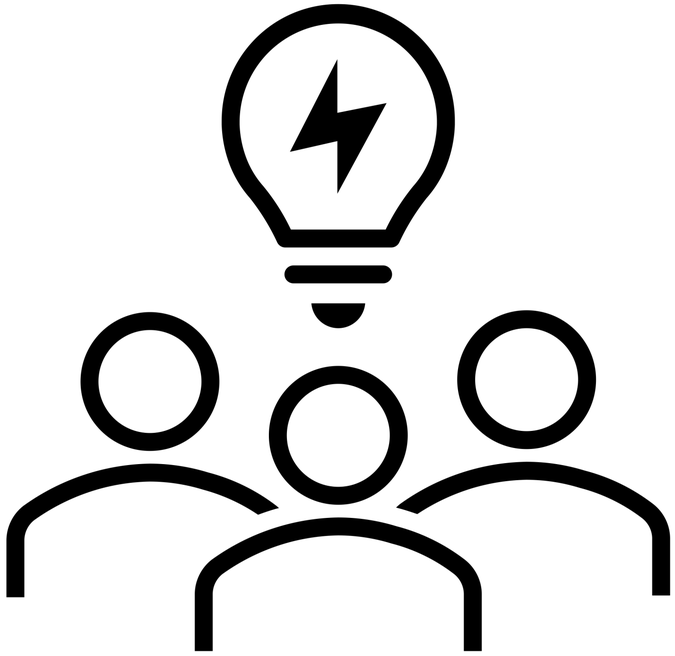
What is Visualization?

What is Visualization?



- Find an **example of a visualization** you have seen recently.
- Add a screenshot of it to the Jamboard here: Real-World-Examples

What is Visualization?



- Look at the examples of visualizations. Talk to your neighbors and **brainstorm a definition for Information Visualization**.

- Add your thoughts to the Jamboard here: [Define-Information-Visualization]

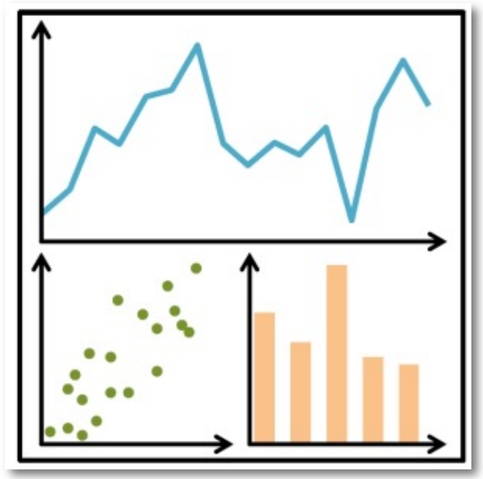
Official Definition

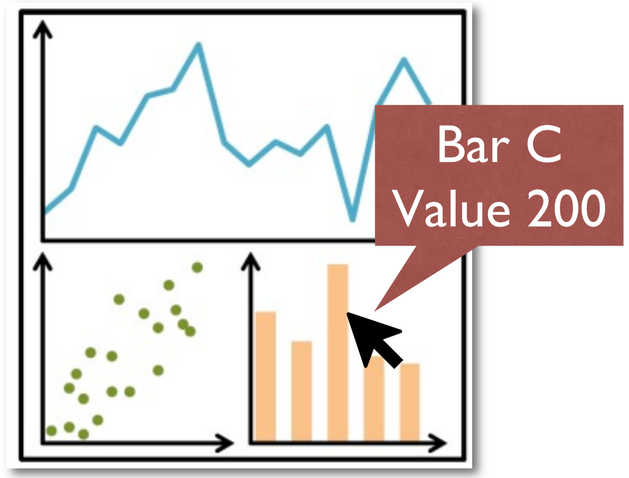
Visualization: the visual representation of data to reinforce human cognition

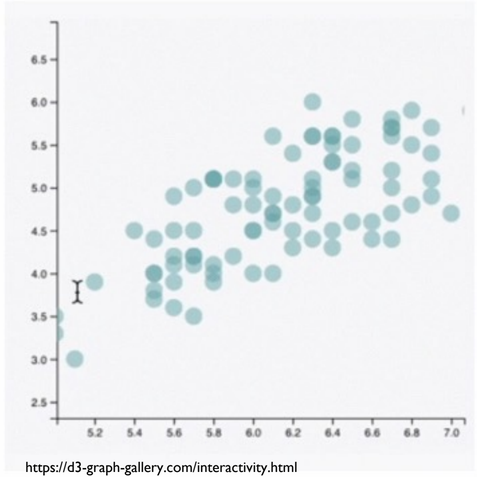
Official Definition

(static or interactive)

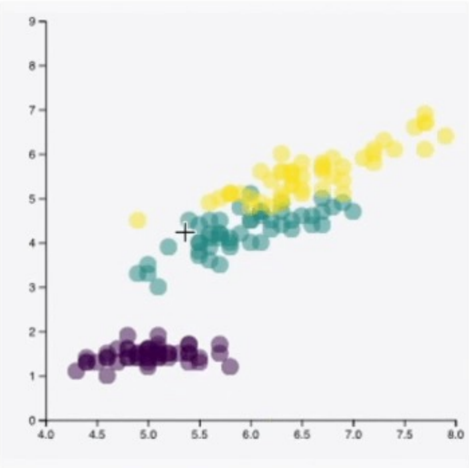
Visualization: the visual representation of data to reinforce human cognition







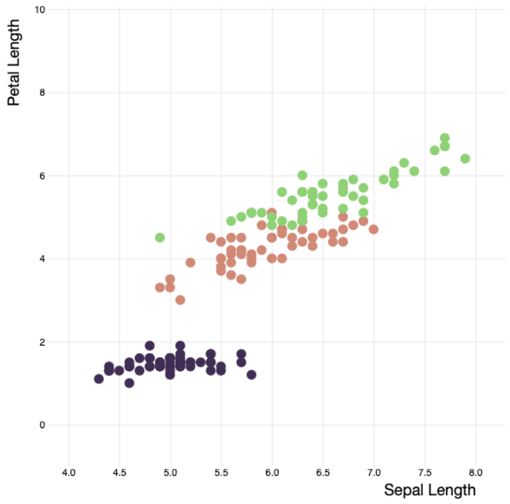
https://d3-graph-gallery.com/interactivity.html

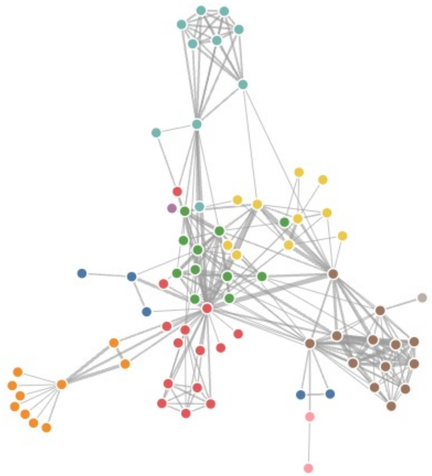


Official Definition

(static or interactive) (abstract or spatial)

Visualization: the visual representation of data to reinforce human cognition





https://d3-graph-gallery.com/

Official Definition

(static or interactive) (abstract or spatial)

Visualization: the visual representation of data to reinforce human cognition

Why Visualization?

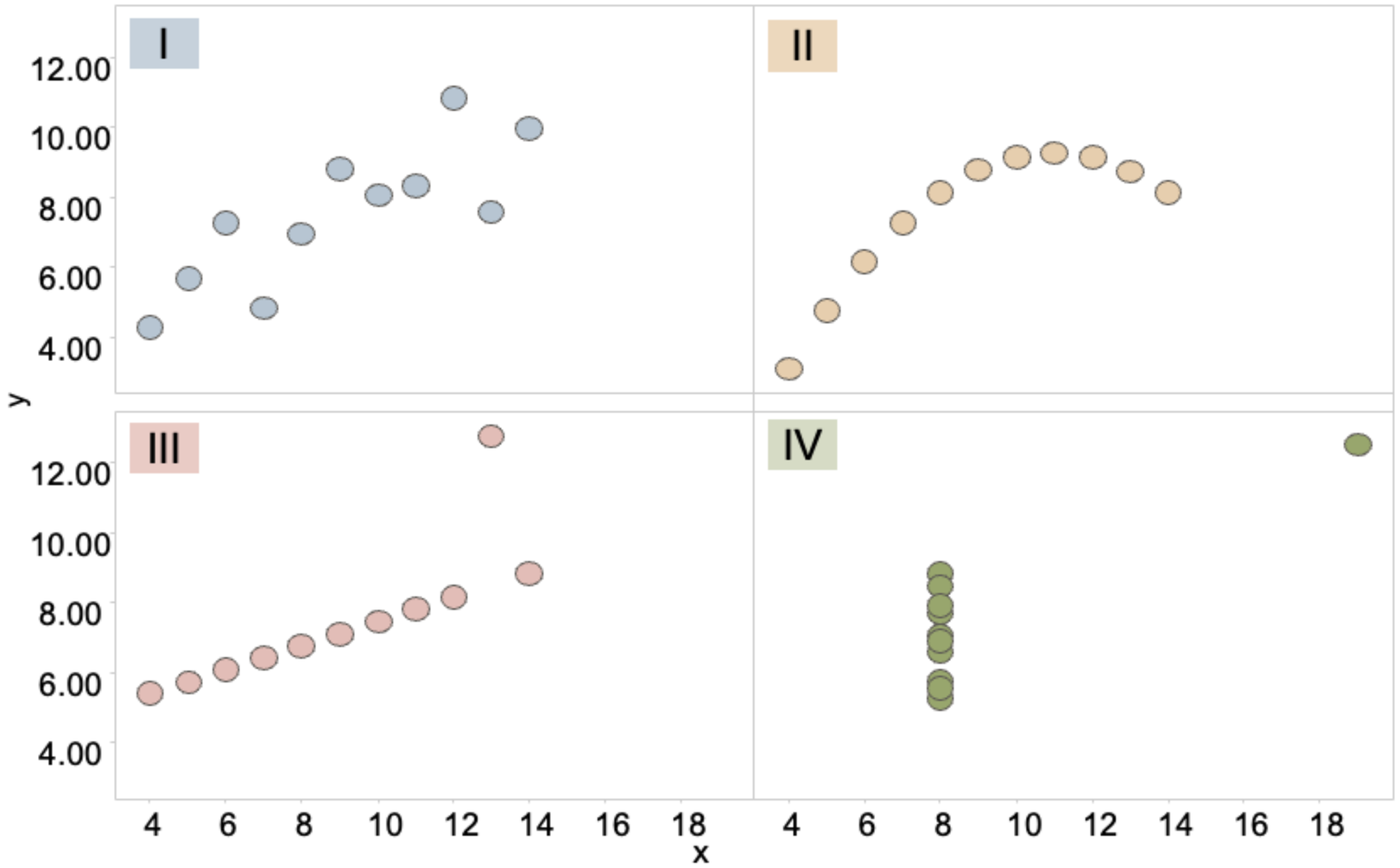
Why visualize data?

I		II		III		IV	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

Why visualize data?

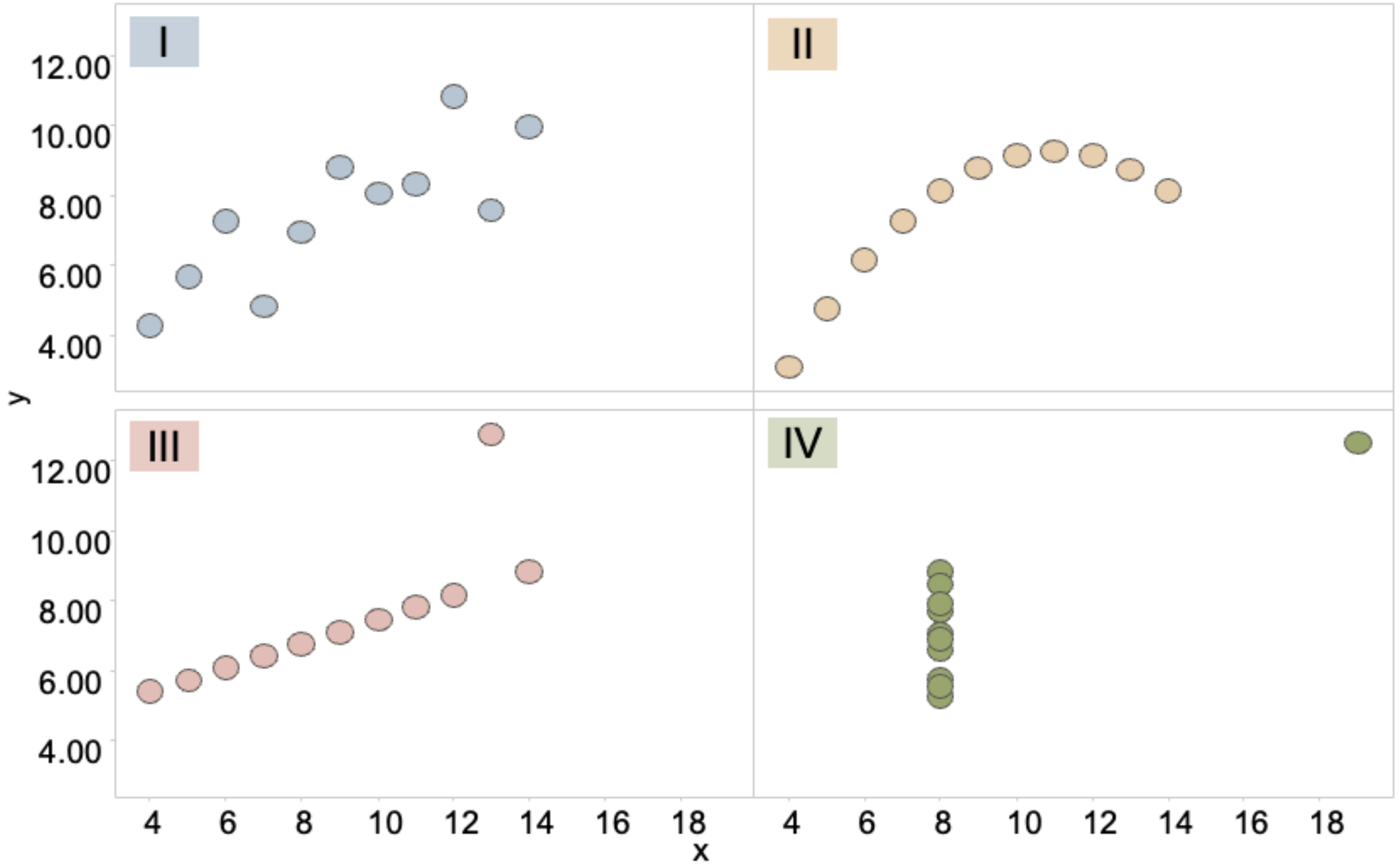
	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
	8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
	13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
	9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
	11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
	14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
	6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
	4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
	12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
	7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
	5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

Why visualize data?



Why visualize data?

Reveal Patterns



Why visualize data?

On the next slide, count the number of “t”s as fast as you can

Why visualize data?

q t f j n i x i g j u n a s b b t g

r k c l b v t x j x z x m x r g k

l x q h m z y w t e y j w n o p

Why visualize data?

Why visualize data?

On the next slide, count the number of “j”s as fast as you can

Why visualize data?

q t f j n i x i g j u n a s b b t g

r k c l b v t x j x z x m x r g k

l x q h m z y w t e y j w n o p

Why visualize data?

Leverage Human Perception

q t f j n i x i g j u n a s b b t g
r k c l b v t x j x z x m x r g k
l x q h m z y w t e y j w n o p

Why visualize data?

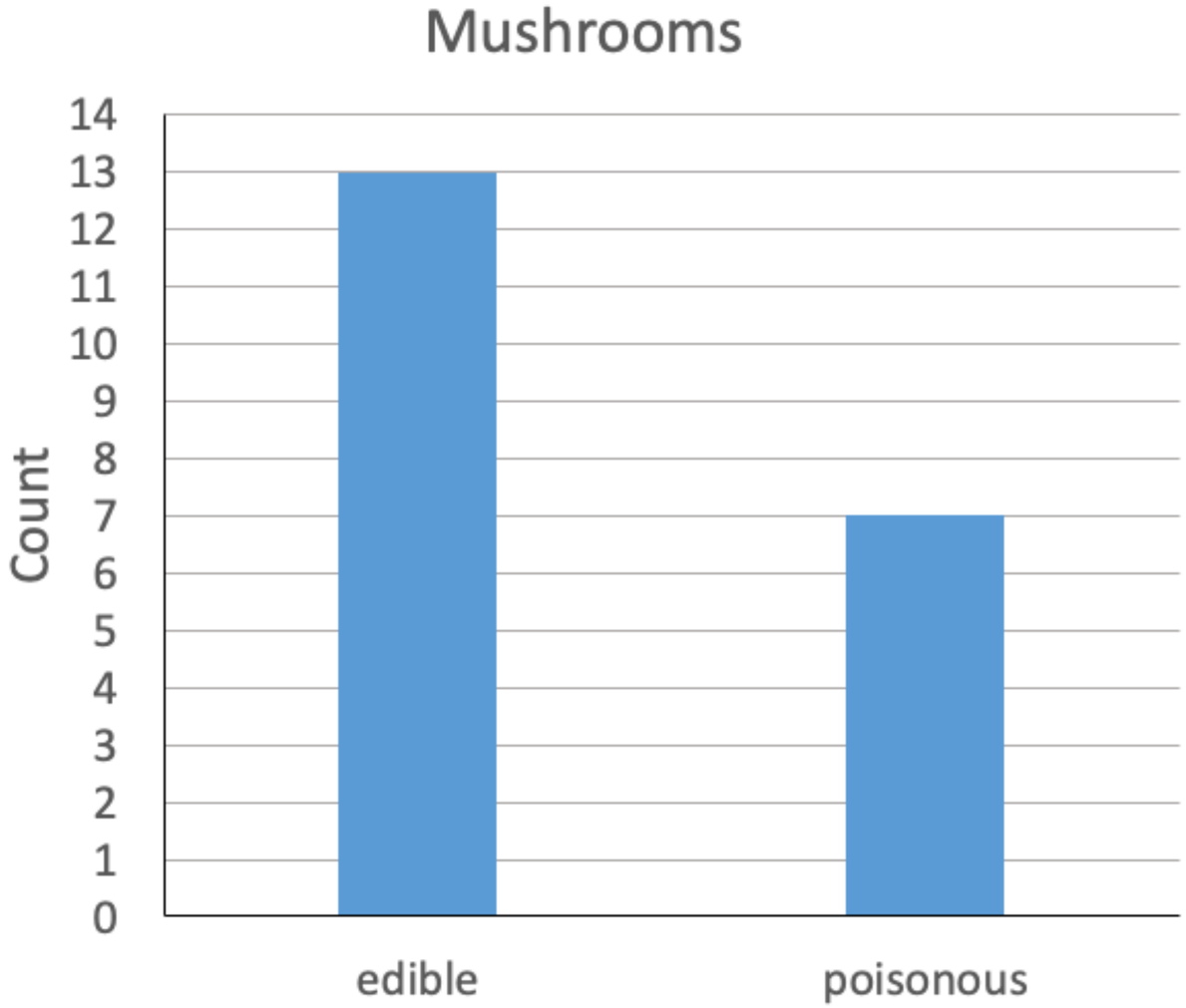
How many mushrooms are poisonous?

class	cap-shape	cap-surface	cap-color	bruises	odor
poisonous	convex	smooth	brown	yes	pungent
edible	convex	smooth	yellow	yes	almond
edible	bell	smooth	white	yes	anise
poisonous	convex	scaly	white	yes	pungent
edible	convex	smooth	gray	no	none
edible	convex	scaly	yellow	yes	almond
edible	bell	smooth	white	yes	almond
edible	bell	scaly	white	yes	anise
poisonous	convex	scaly	white	yes	pungent
edible	bell	smooth	yellow	yes	almond
edible	convex	scaly	yellow	yes	anise
edible	convex	scaly	yellow	yes	almond
edible	bell	smooth	yellow	yes	almond
poisonous	convex	scaly	white	yes	pungent
edible	convex	fibrous	brown	no	none
edible	sunken	fibrous	gray	no	none
edible	flat	fibrous	white	no	none
poisonous	convex	smooth	brown	yes	pungent
poisonous	convex	scaly	white	yes	pungent
poisonous	convex	smooth	brown	yes	pungent

https://domohelp.domo.com/hc/en-us/articles/360043931814-Fun-Sample-DataSets

Why visualize data?

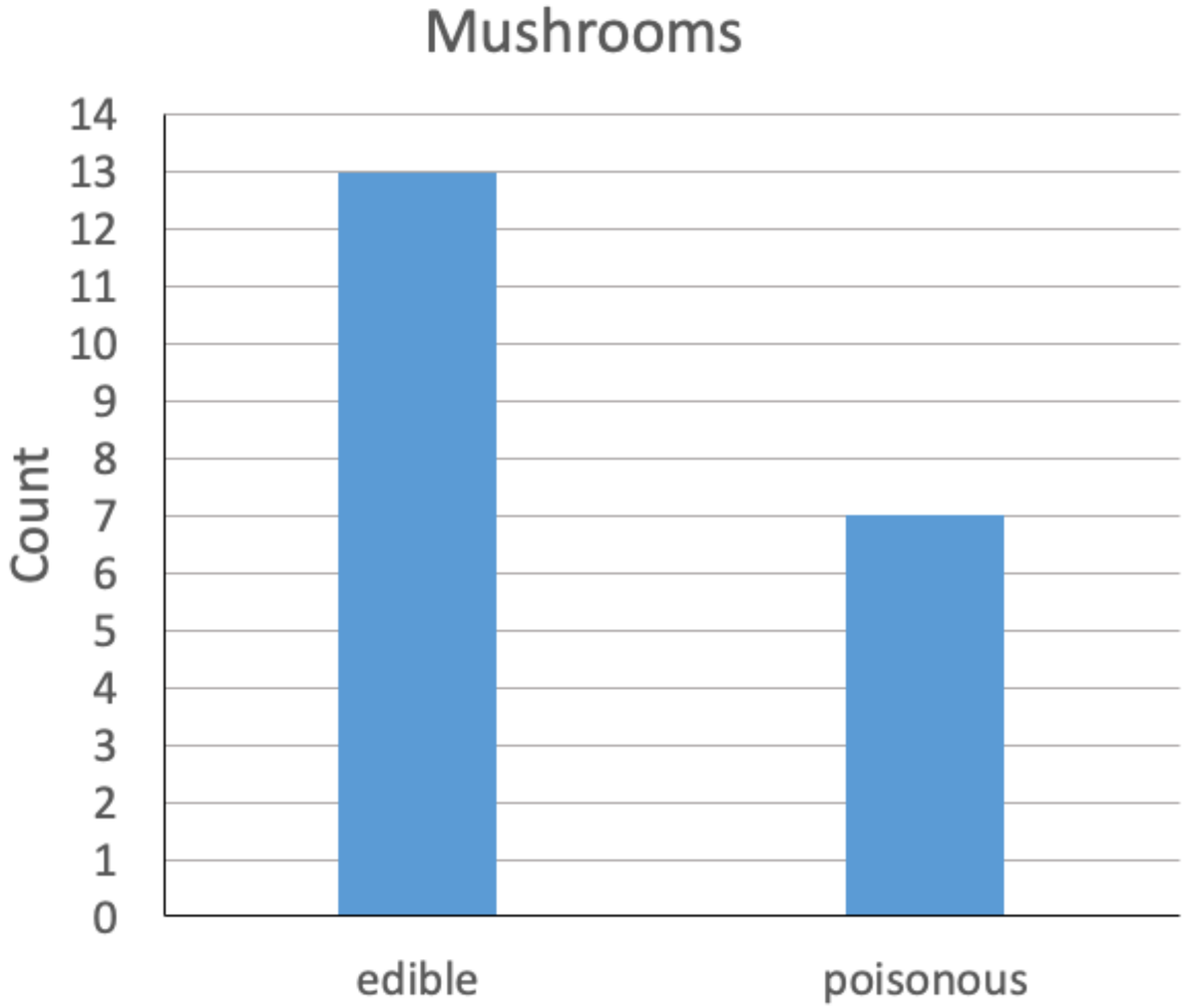
How many mushrooms are poisonous?



https://domohelp.domo.com/hc/en-us/articles/360043931814-Fun-Sample-DataSets

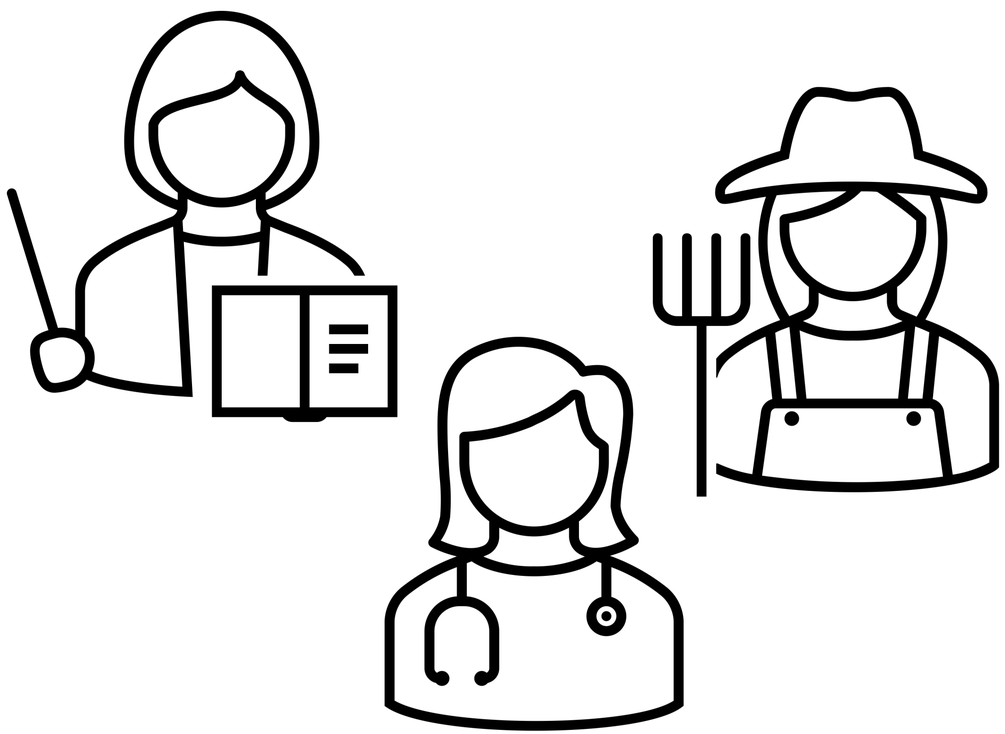
Why visualize data?

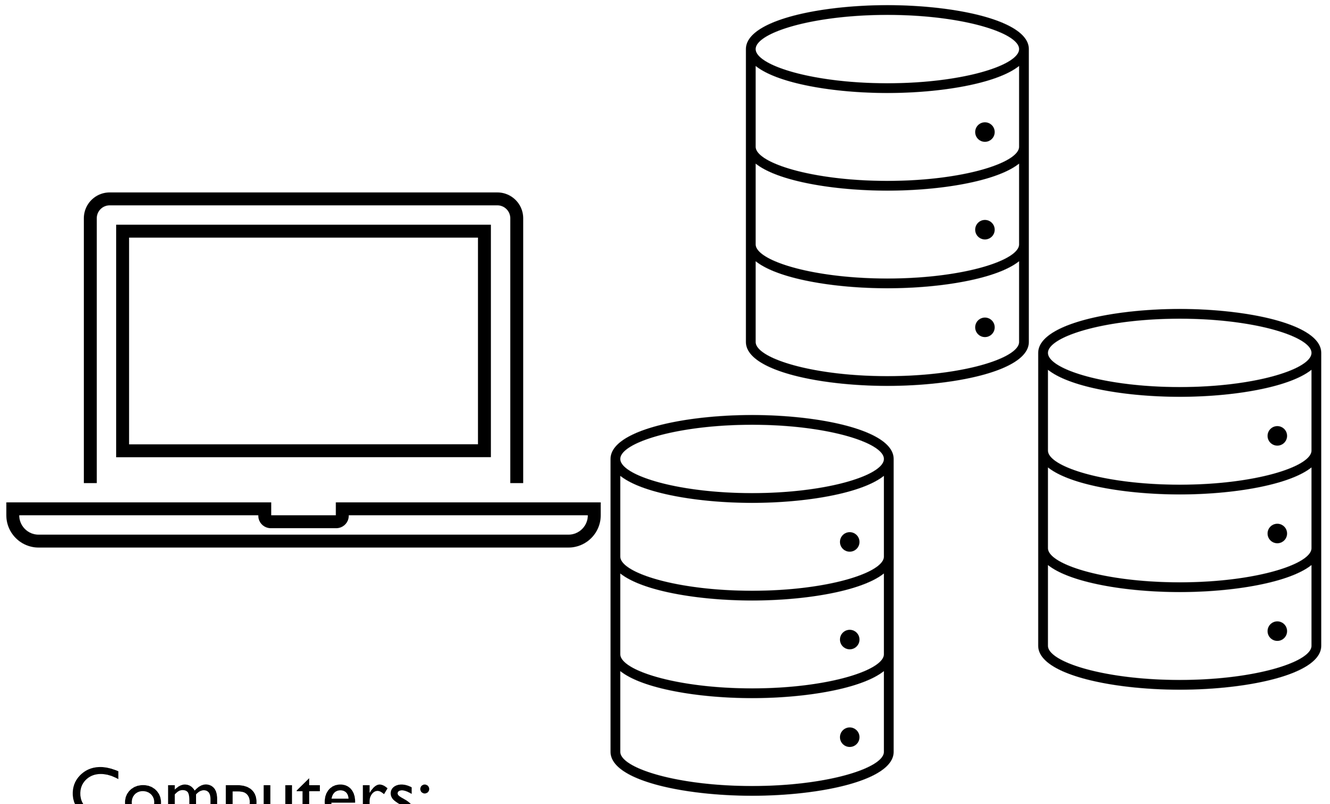
Reduce Cognitive Load



https://domohelp.domo.com/hc/en-us/articles/360043931814-Fun-Sample-DataSets

Human - Computer Collaboration





Humans:

- Have subject matter expertise for guiding analyses and using output

Computers:

- Can speed up large data analyses
- Enable record keeping

Can visualization actually make a difference?

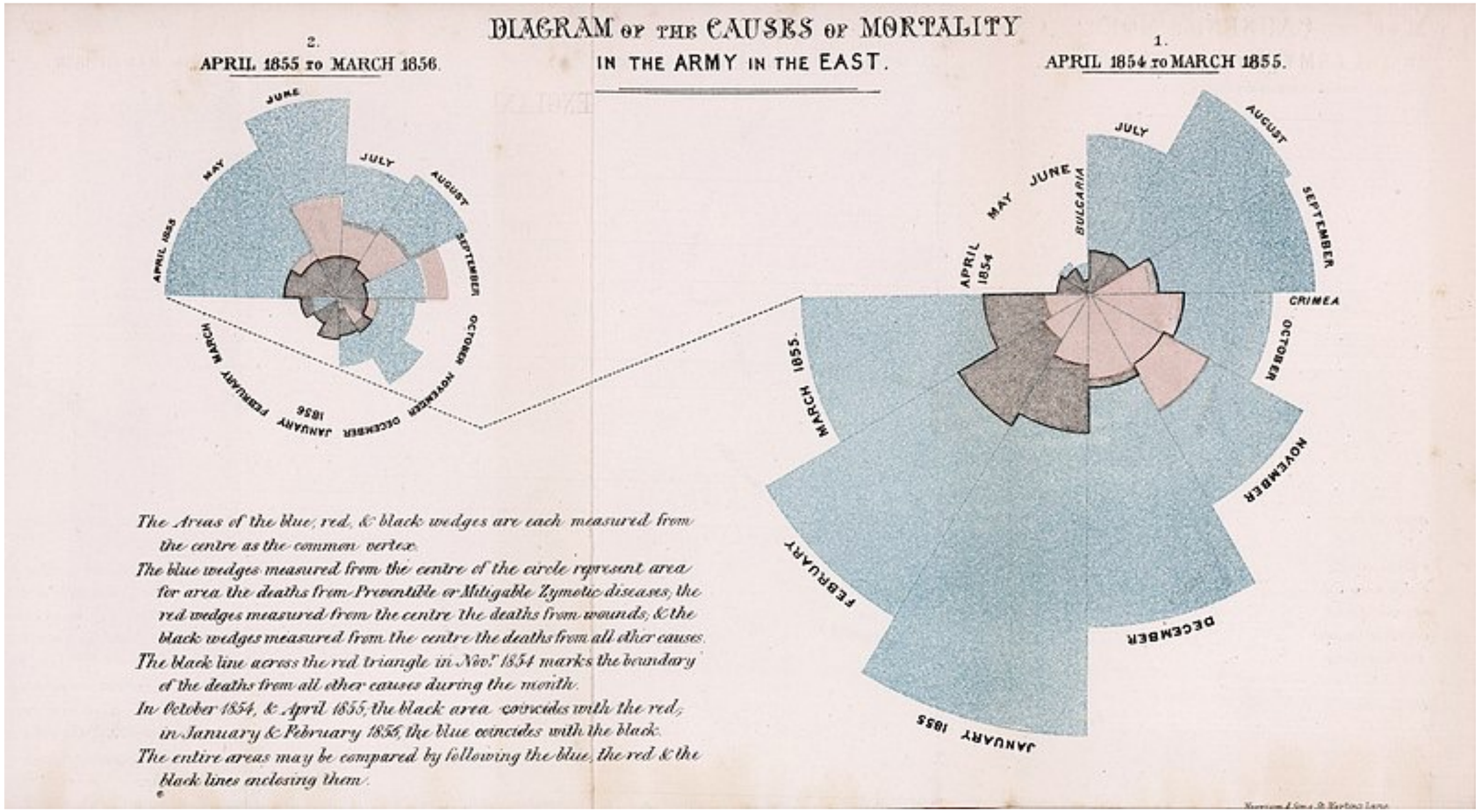
Does it really work?

John Snow's Cholera Map

https://www.wired.com/2009/09/0908london-cholera-pump/

Does it really work?

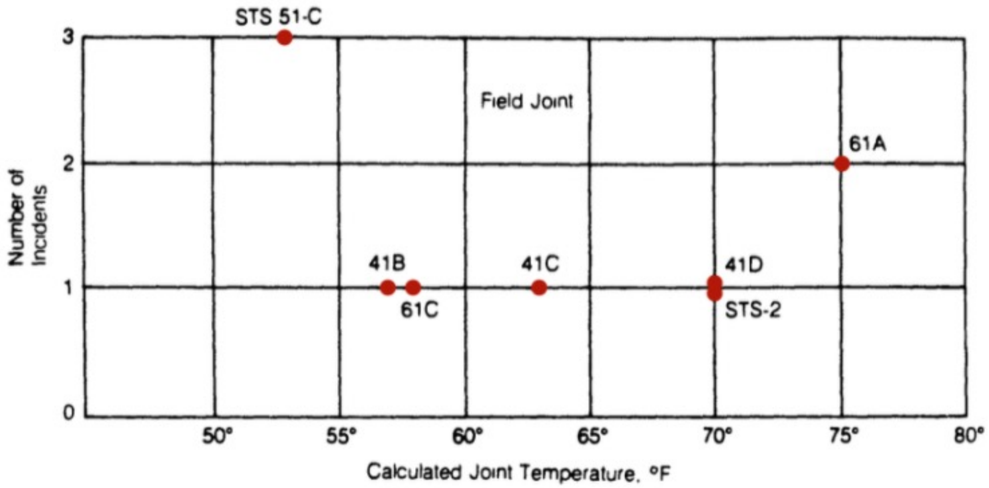
Florence Nightingale's Rose



https://commons.wikimedia.org/wiki/File:Nightingale-mortality.jpg

Does it really work?

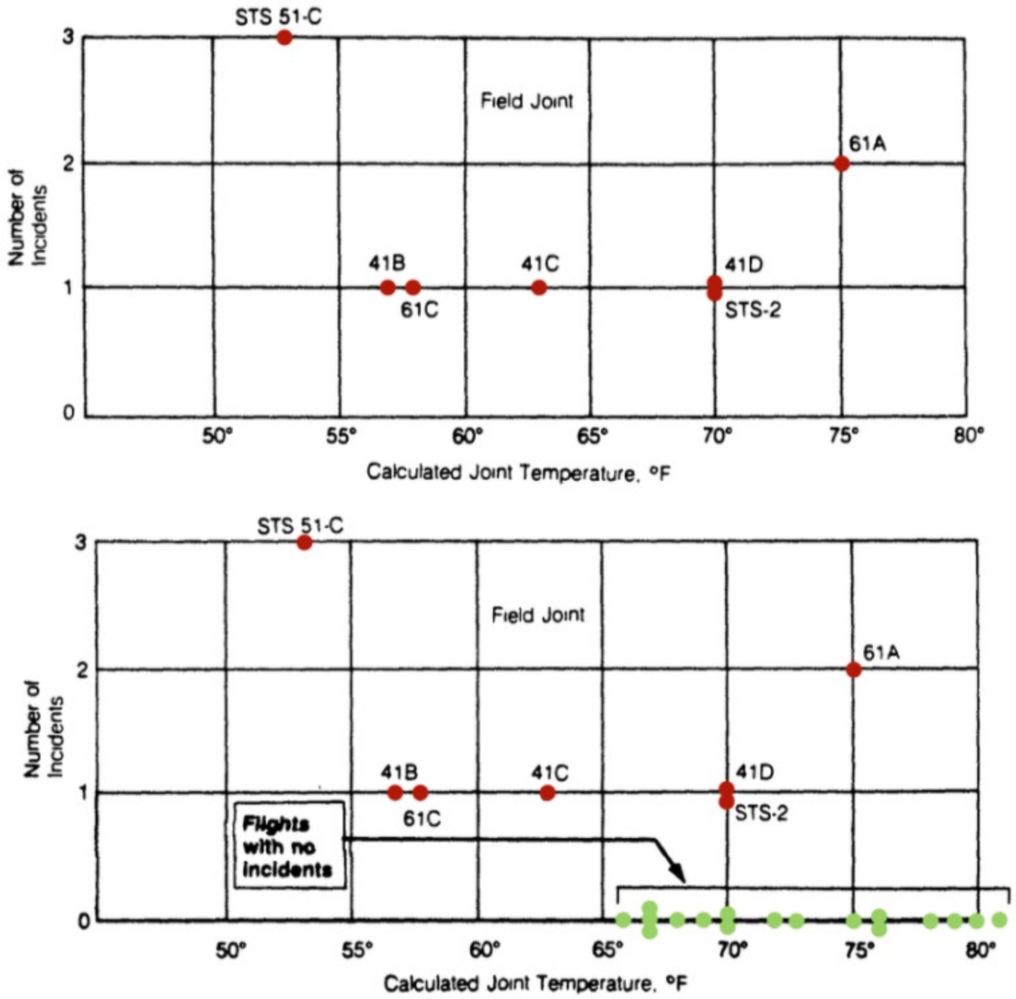
Challenger Shuttle



https://studentwork.prattsi.org/infovis/labs/a-history-of-dangerously-misleading-data-visualization/

Does it really work?

Challenger Shuttle

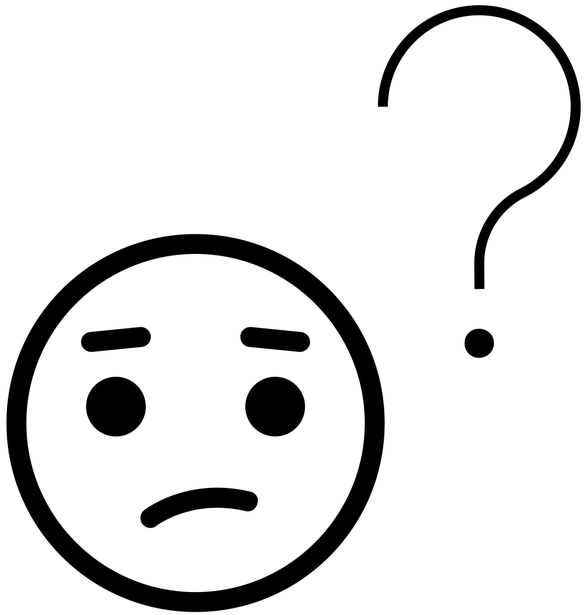


https://studentwork.prattsi.org/infovis/labs/a-history-of-dangerously-misleading-data-visualization/

Let’s take a break! Stretch, go for a walk, be social ☺
Be back here in 10 mins.

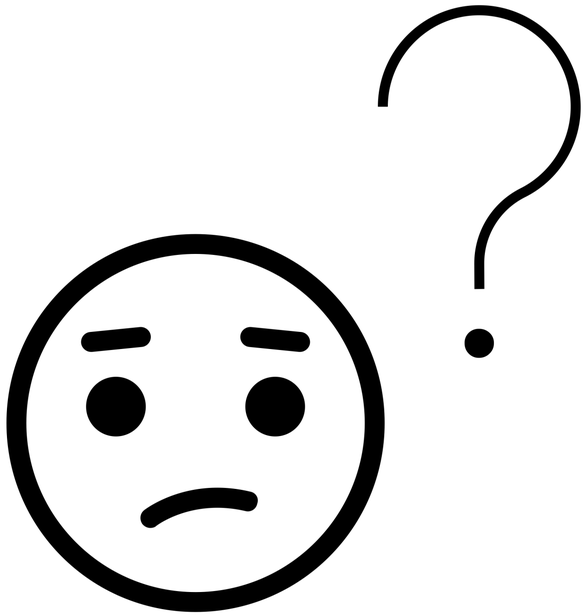
Intro to Git

Introduction to Git



What is Git?

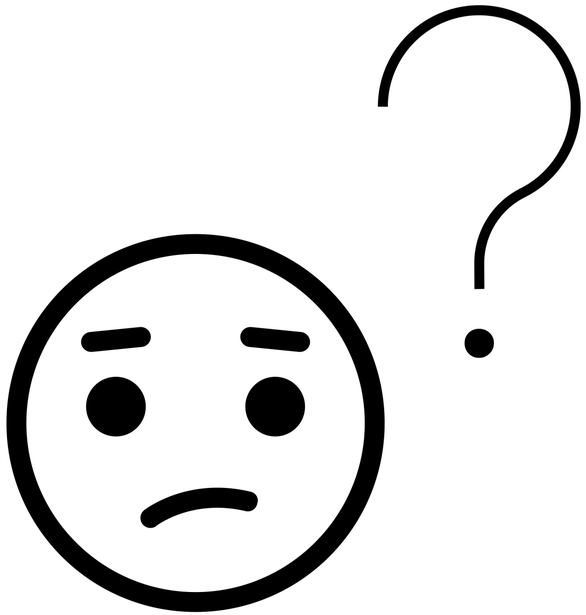
Introduction to Git



What is Git?

A Distributed Version Control System

Introduction to Git

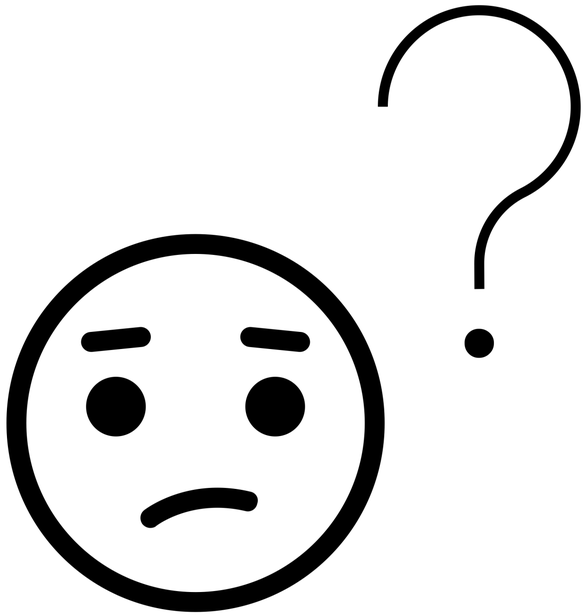


What is Git?

A Distributed Version Control System

stores content and tracks changes

Introduction to Git



What is Git?

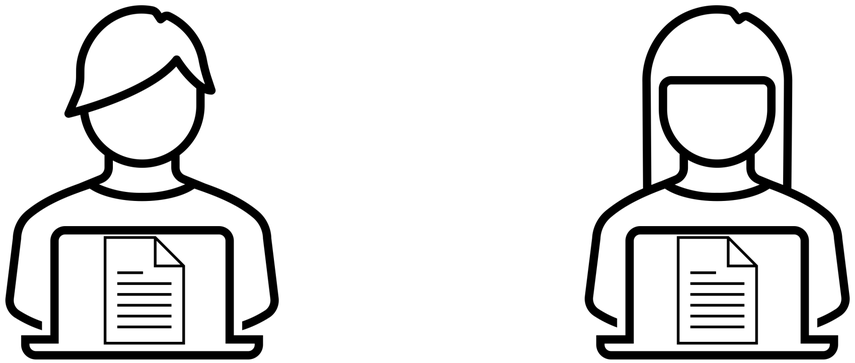
A Distributed Version Control System

stores content and tracks changes

content stored in multiple places

Introduction to Git

Remote = copy on the server



Local = copy on your computer

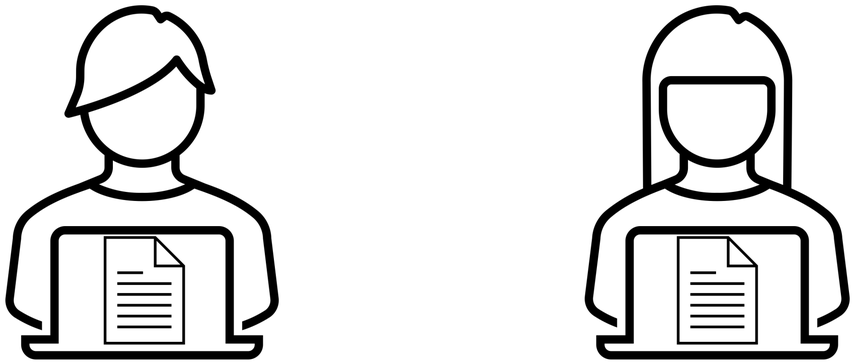
What is Git?

A Distributed Version Control System

- stores content and tracks changes
- content stored in multiple places

Introduction to Git

Remote = copy on the server



Local = copy on your computer

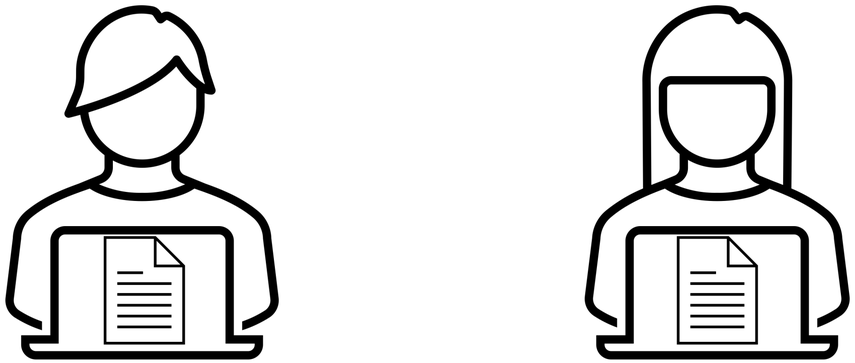
Why use Git?

To keep backups of your work

To make collaboration easier

Introduction to Git

Remote = copy on the server



Local = copy on your computer

Why use Git?

To keep backups of your work

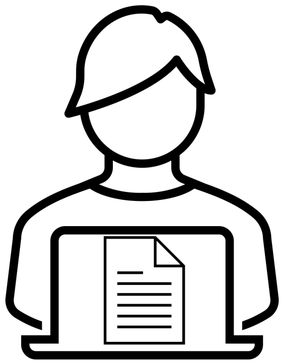
To make collaboration easier

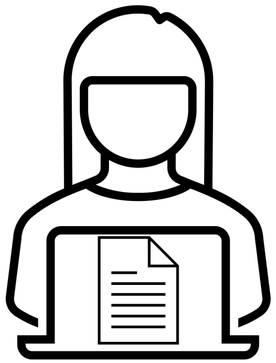
Ex. Suppose you finished part 1 of your homework. You move on to part 2 but part way through realize you broke part 1 and cannot figure out how.

Luckily, you backed up your work to the remote server so you can compare what you have now to what you had when part 1 was working and fix the issue.

Introduction to Git

Remote = copy on the server





Local = copy on your computer

Why use Git?

To keep backups of your work

To make collaboration easier

Ex. Suppose you and your partner are working on your final project. They are working on vis 1 and you are working on vis 2. You will combine your code later.

When you are each finished and ready to combine your two copies of code, you can use Git to automatically find conflicts and make the process smoother.

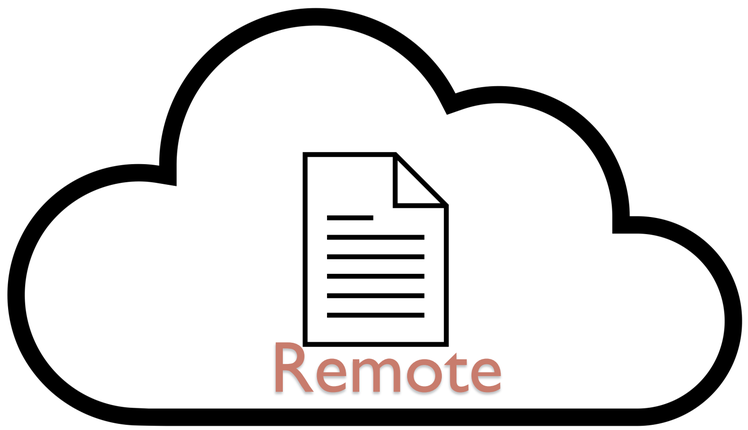
Introduction to Git

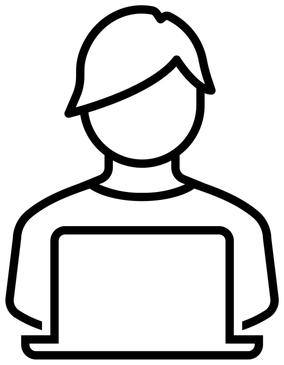
How to use Git?

You can see the remote version of your code through the browser with GitHub.

Your local version is what you see in the file system on your computer.

Introduction to Git

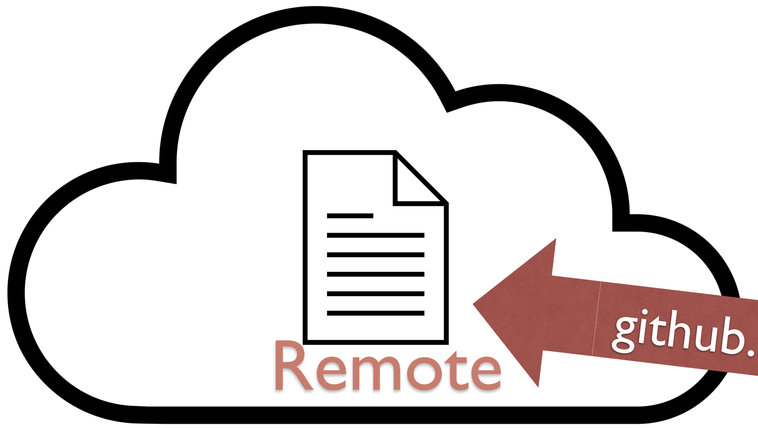




How to use Git?

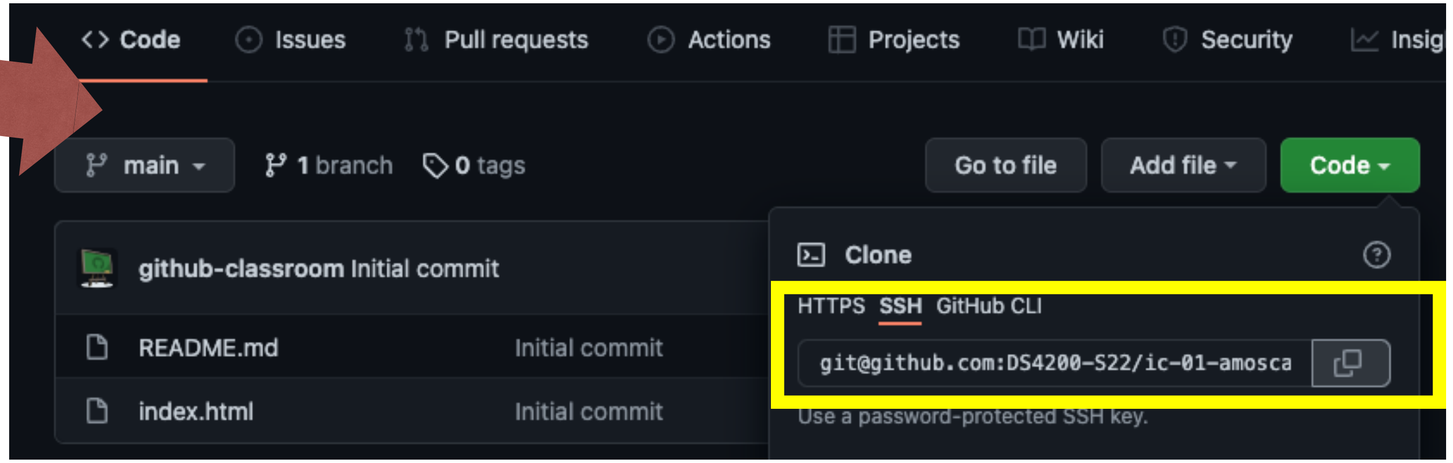
To get a local version of remote files for the first time, you clone the remote version (which is called a repository or repo).

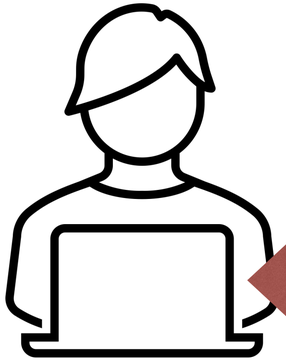
Introduction to Git



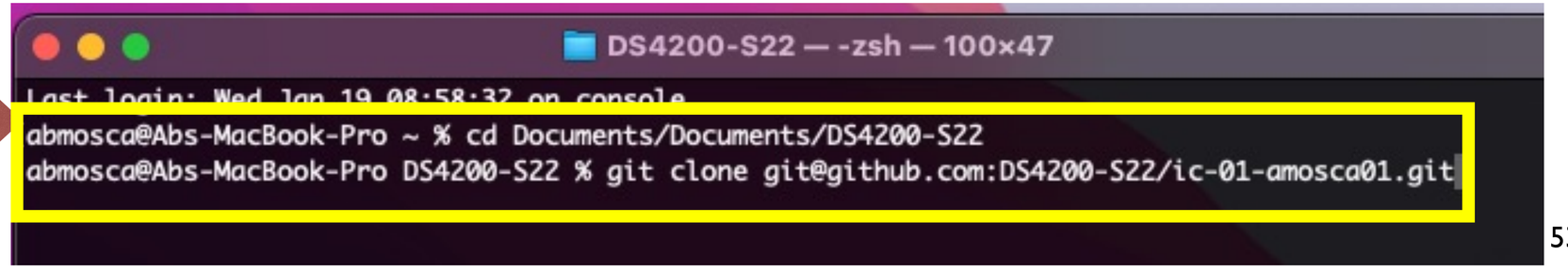
How to use Git?

To get a local version of remote files for the first time, you clone the remote version (which is called a repository or repo).

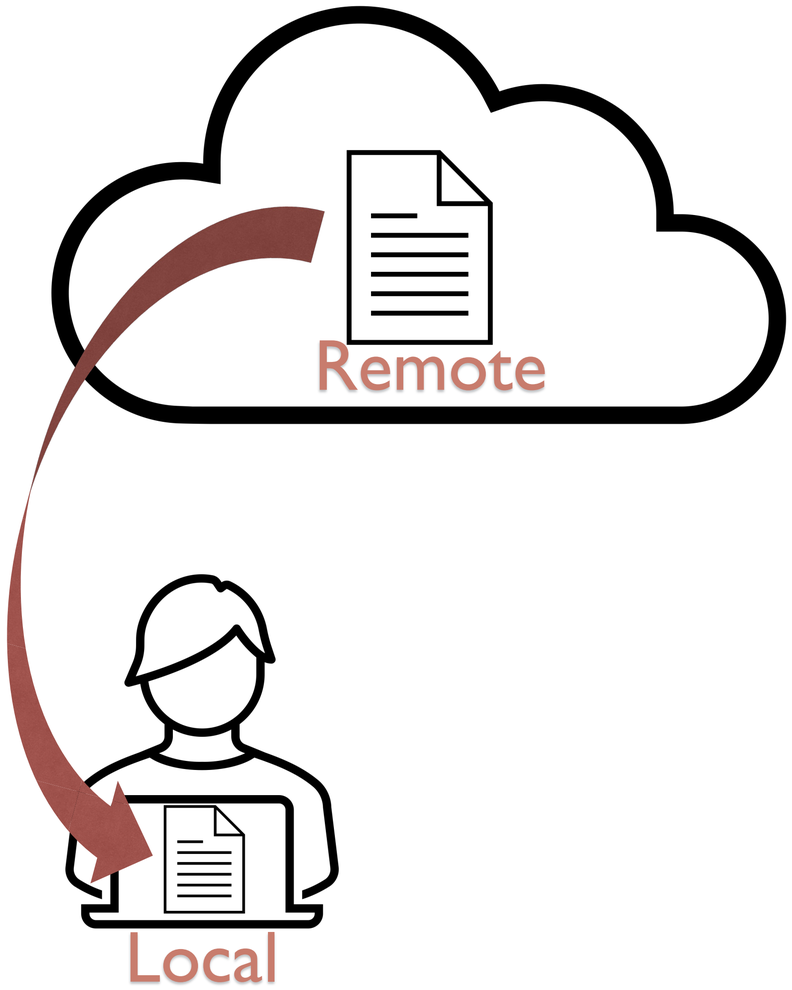




command line

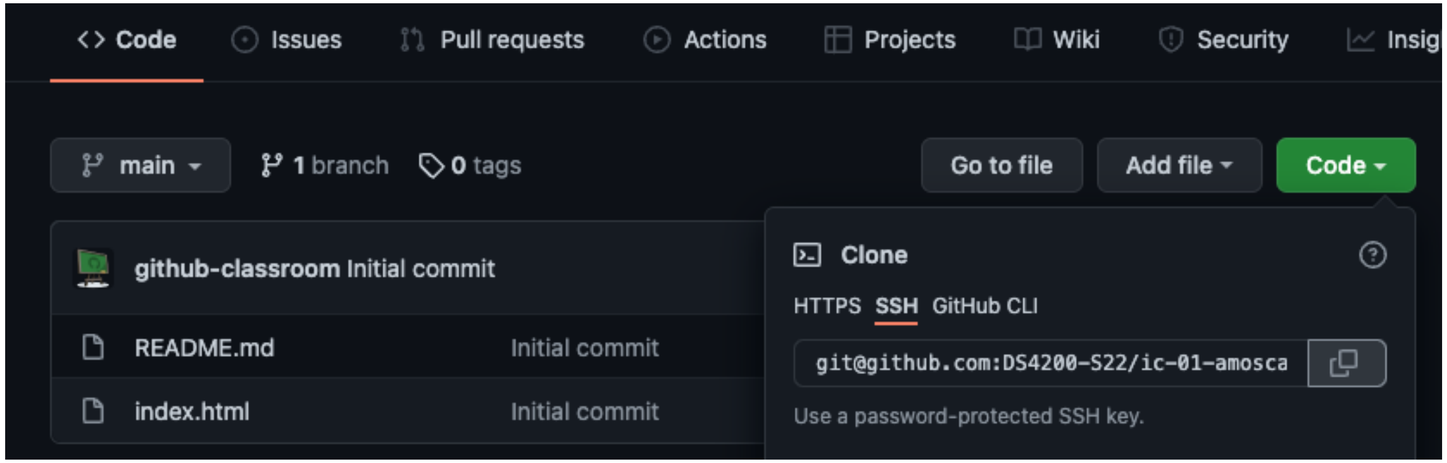


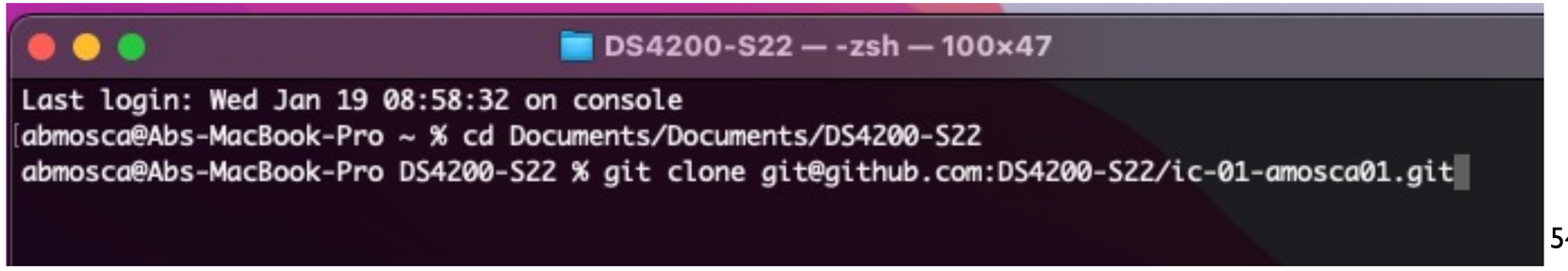
Introduction to Git



How to use Git?

To get a local version of remote files for the first time, you clone the remote version (which is called a repository or repo).





Introduction to Git

How to use Git?

To add work from your local repo to the remote repo, add and commit your local changes and push them to the remote repo.

Introduction to Git

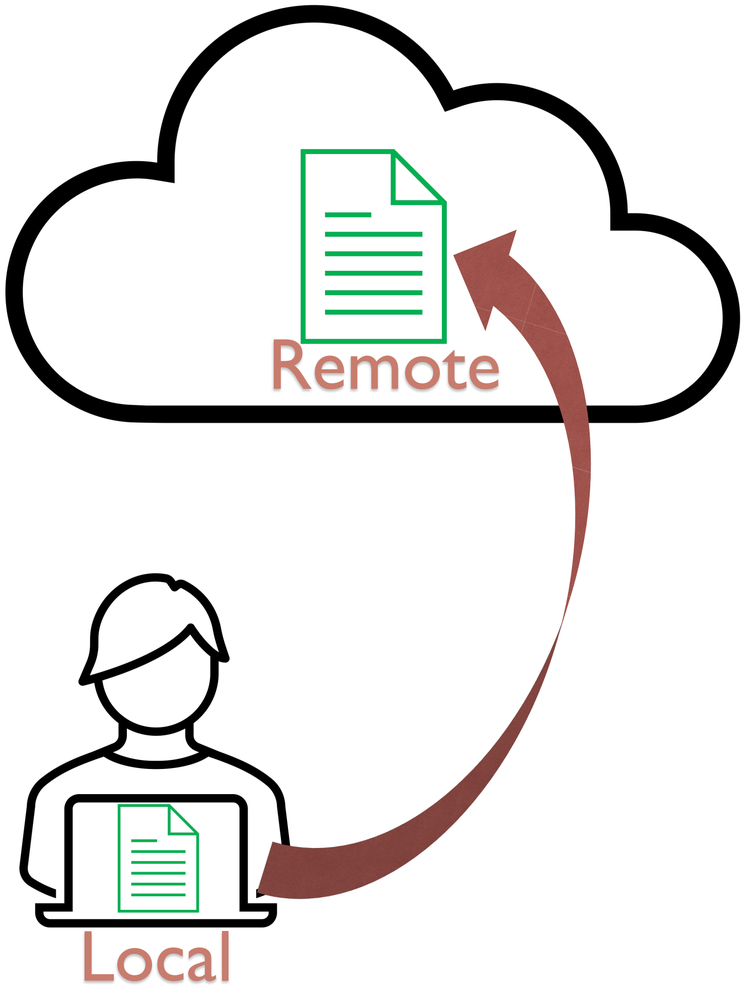
command line

How to use Git?

To add work from your local repo to the remote repo, add and commit your local changes and push them to the remote repo.

```
ic-01-amosca01 — -zsh — 100×47
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git add .
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git commit -m "part 1 done"
[main ac03563] part 1 done
 1 file changed, 1 insertion(+), 1 deletion(-)
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git push
Enumerating objects: 5, done.
Counting objects: 100% (5/5), done.
Delta compression using up to 8 threads
Compressing objects: 100% (3/3), done.
Writing objects: 100% (3/3), 306 bytes | 306.00 KiB/s, done.
Total 3 (delta 1), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (1/1), completed with 1 local object.
To github.com:DS4200-S22/ic-01-amosca01.git
   cf7b50c..ac03563  main -> main
abmosca@Abs-MacBook-Pro ic-01-amosca01 %
```

Introduction to Git



How to use Git?

To add work from your local repo to the remote repo, add and commit your local changes and push them to the remote repo.

```
ic-01-amosca01 — -zsh — 100×47
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git add .
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git commit -m "part 1 done"
[main ac03563] part 1 done
 1 file changed, 1 insertion(+), 1 deletion(-)
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git push
Enumerating objects: 5, done.
Counting objects: 100% (5/5), done.
Delta compression using up to 8 threads
Compressing objects: 100% (3/3), done.
Writing objects: 100% (3/3), 306 bytes | 306.00 KiB/s, done.
Total 3 (delta 1), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (1/1), completed with 1 local object.
To github.com:DS4200-S22/ic-01-amosca01.git
   cf7b50c..ac03563  main -> main
abmosca@Abs-MacBook-Pro ic-01-amosca01 %
```

GitHub in This Class

How will we use Git?

- Starter code for assignments will be distributed through GitHub Classroom
- You will **clone** a repo to get starter code
- To do the assignment, **work locally** on your computer
- Archive backups and **submit** your finished assignment by:

  → **Pushing to GitHub**

  → **Setting up a GitHub Page**

In-class Activity ic-01

Summary

Today we:

- Introductions
- Course Overview
- What is Visualization? Why Visualization?
- Intro to GitHub

hw-00 is OUT today and DUE before next class.
ic-01 is DUE today.