

Lecture I: Introduction, Course Overview, Git

DS 4200
FALL 2022

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

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.

*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>

Teaching Staff Introductions

TA's:



Vedanshi Shah
(she/her)



Monroe Chung
(any pronouns)

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

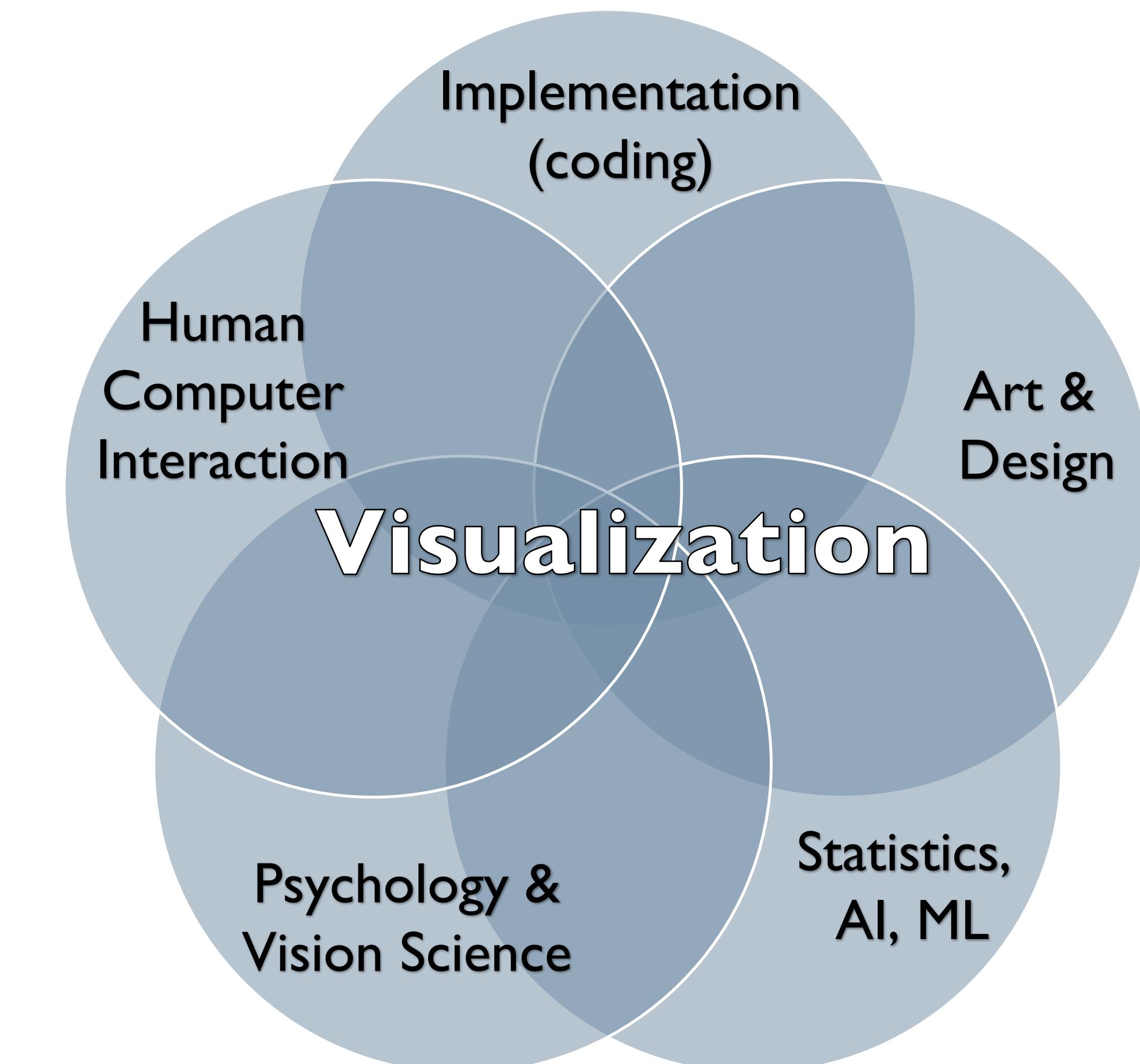
Book a time with Kate here: <https://subjectguides.lib.neu.edu/dataviz/consults>

Course Overview

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



Course Overview

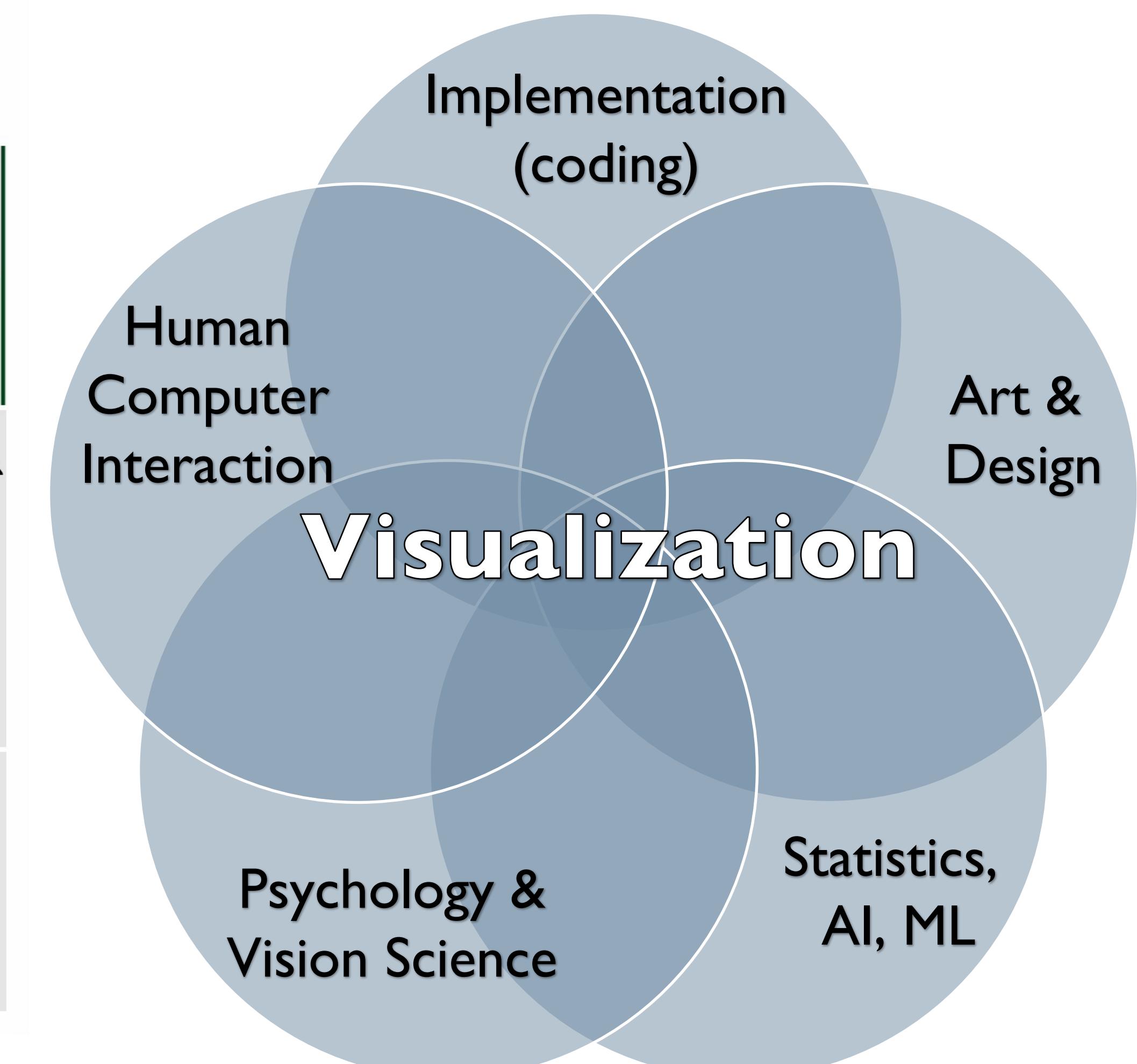
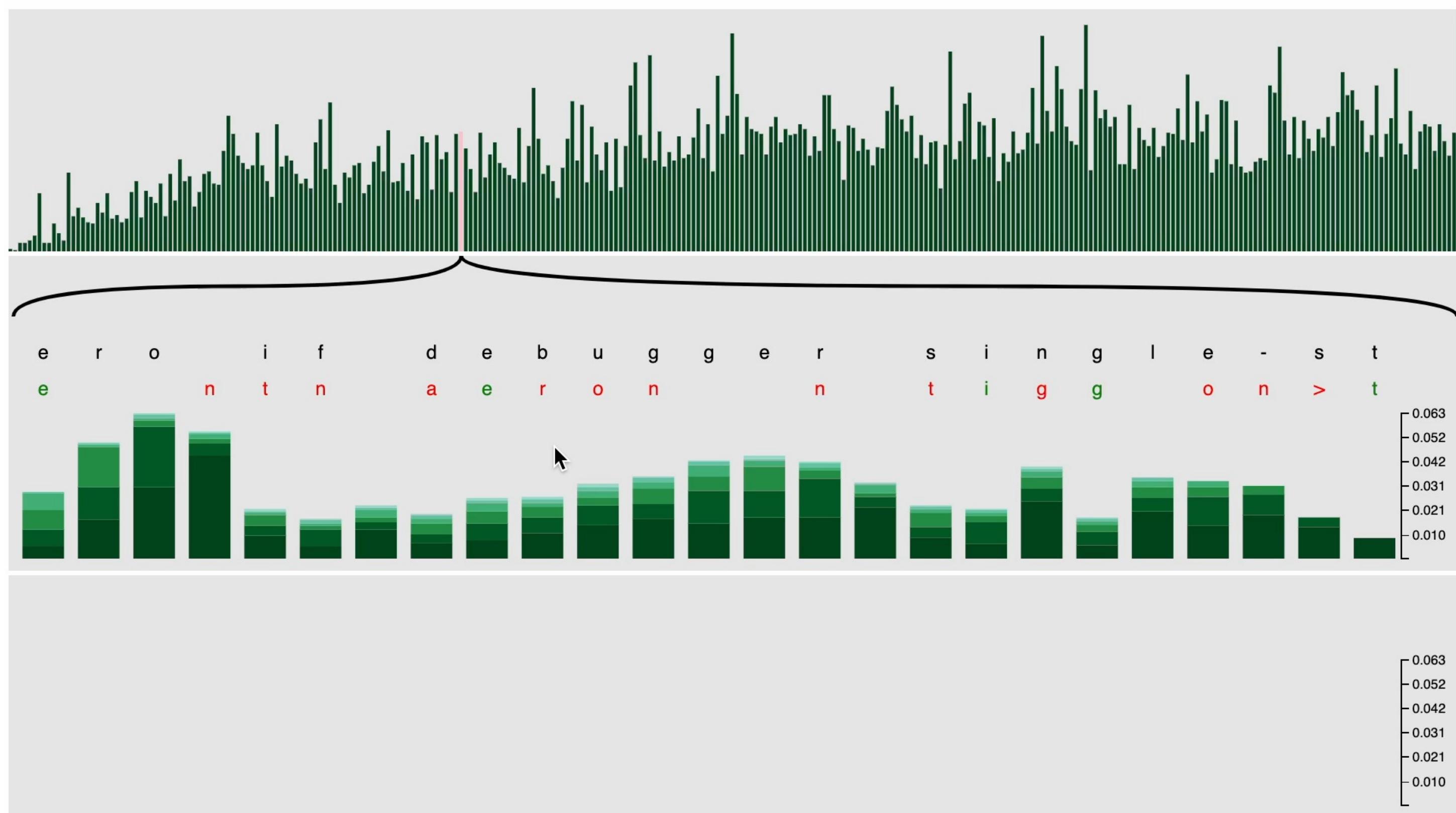


Course Overview

RNNbow

Visualizing Learning via Backpropagation Gradients in Recurrent Neural Networks

Batch Number: 93
Training Cells: 232500 - 232524
Maximum Gradient: 0.0628

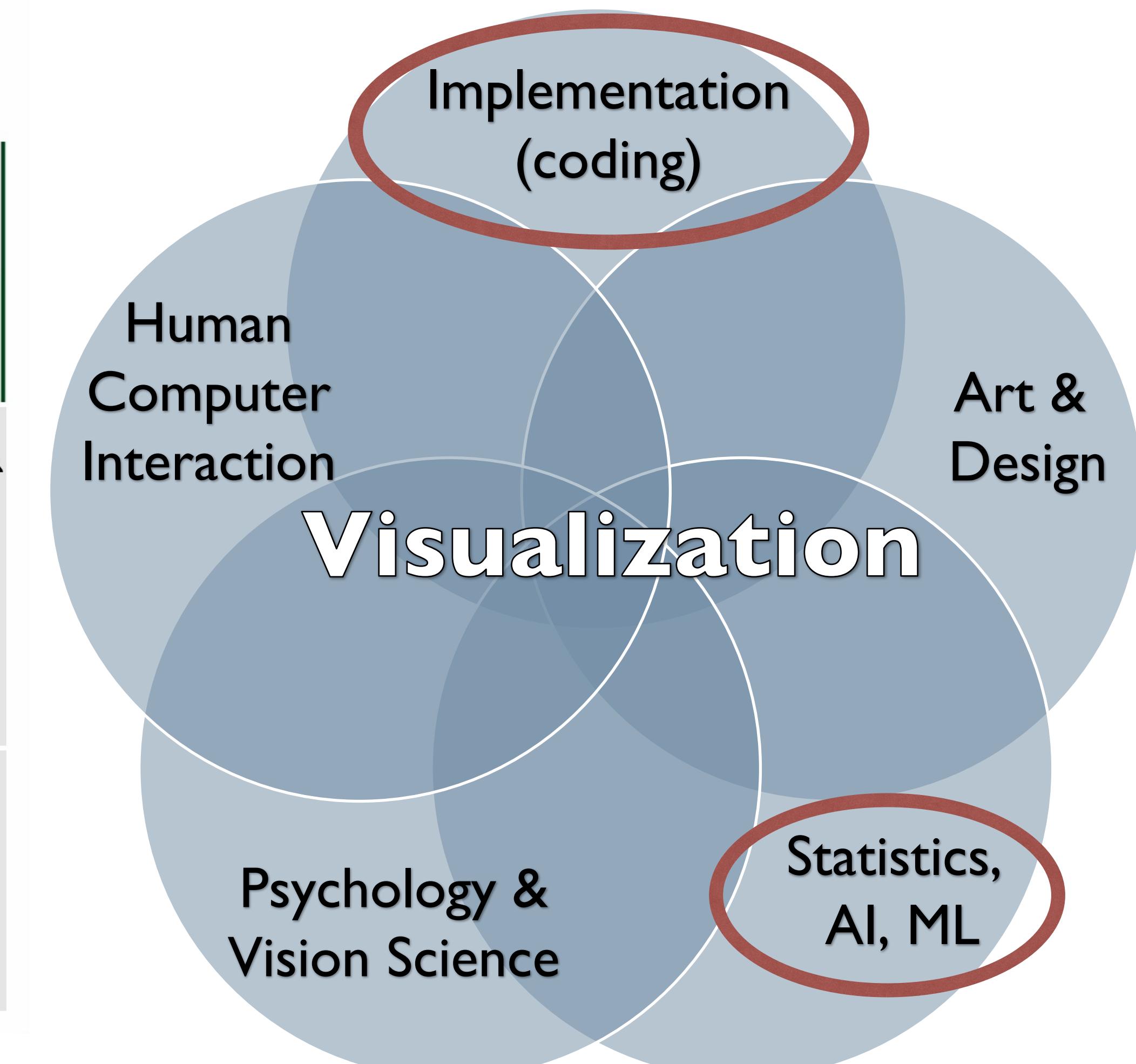
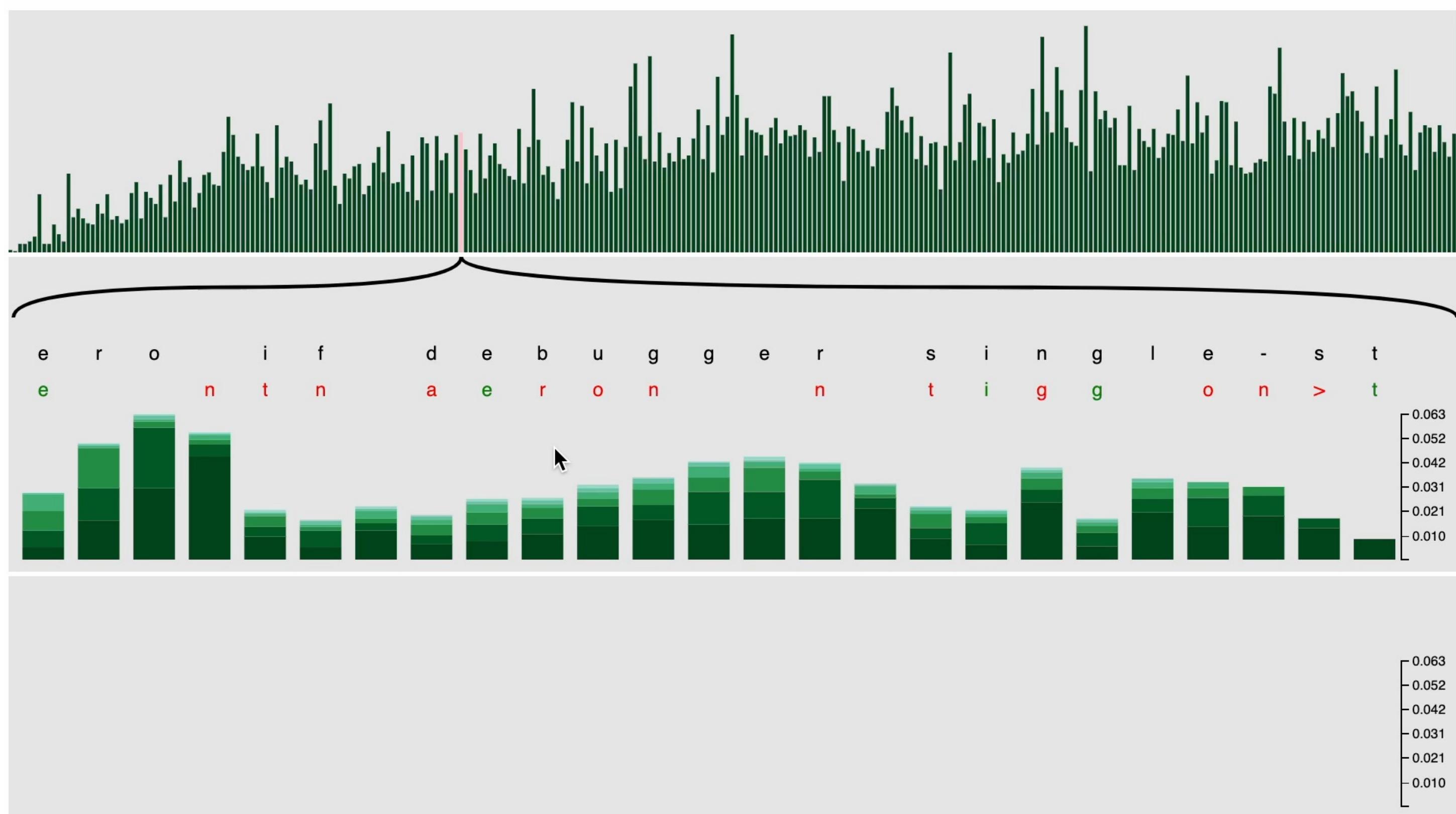


Course Overview

RNNbow

Visualizing Learning via Backpropagation Gradients in Recurrent Neural Networks

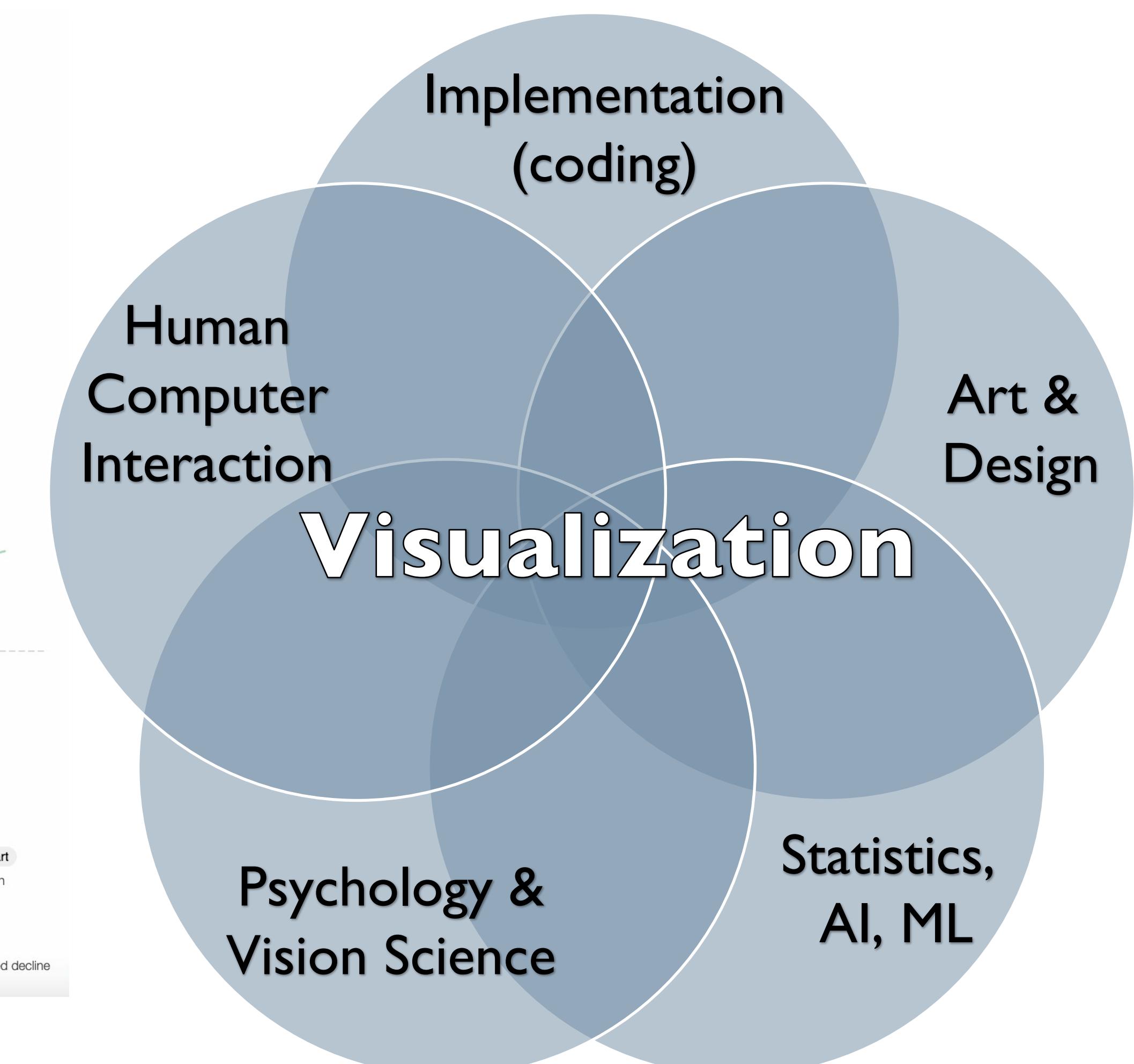
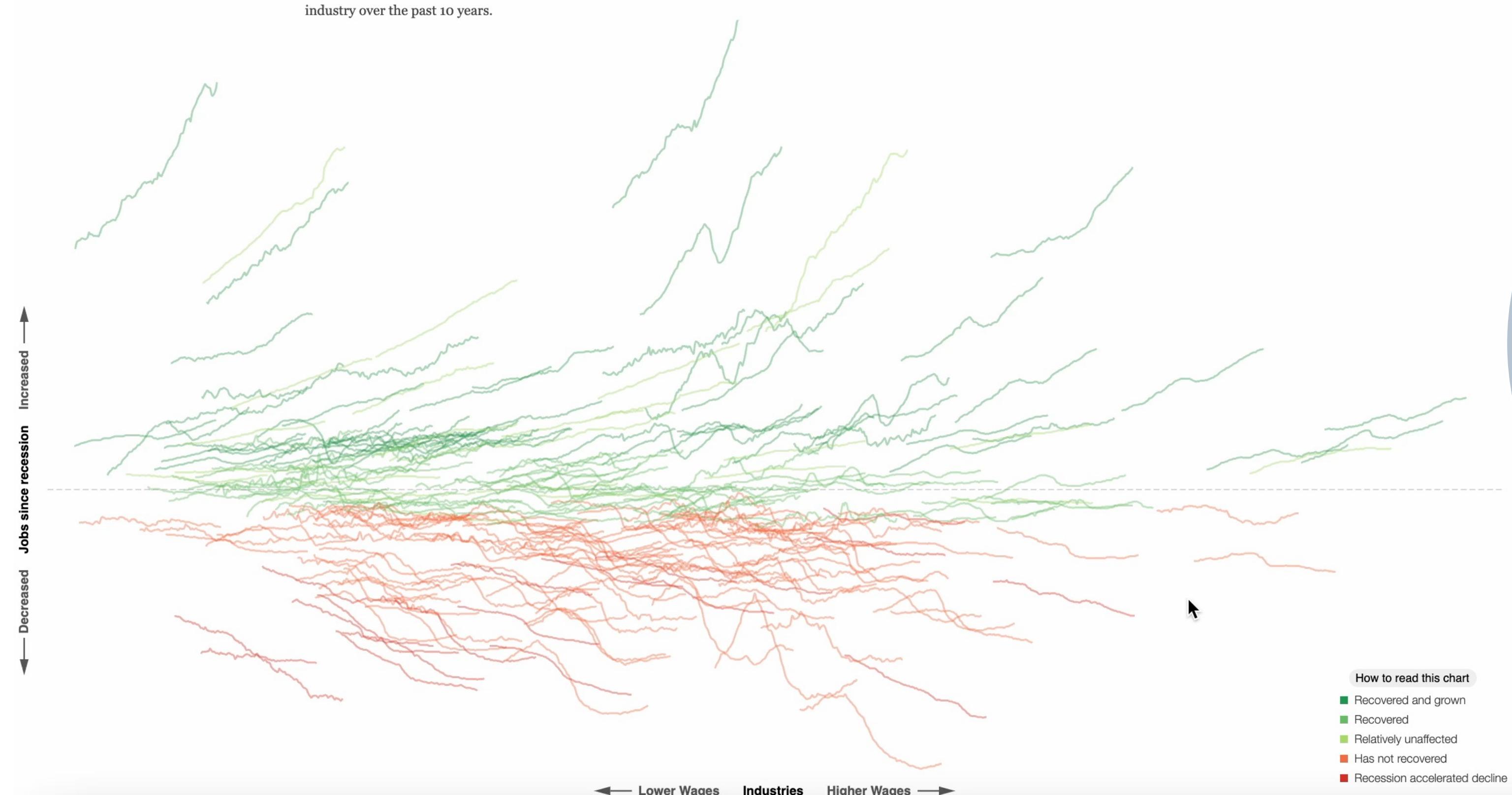
Batch Number: 93
Training Cells: 232500 - 232524
Maximum Gradient: 0.0628



Course Overview

Analyzing trends in the economy

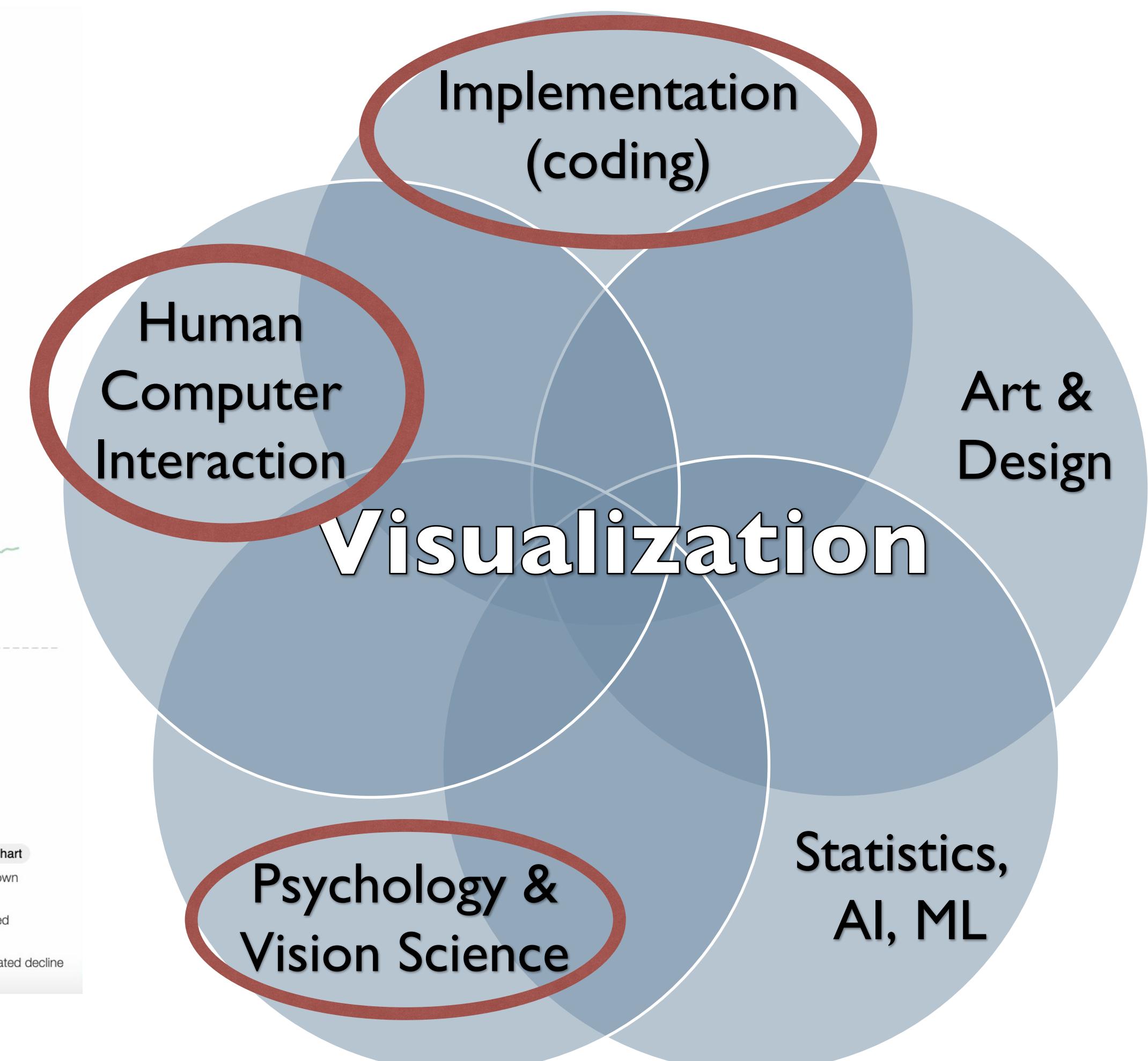
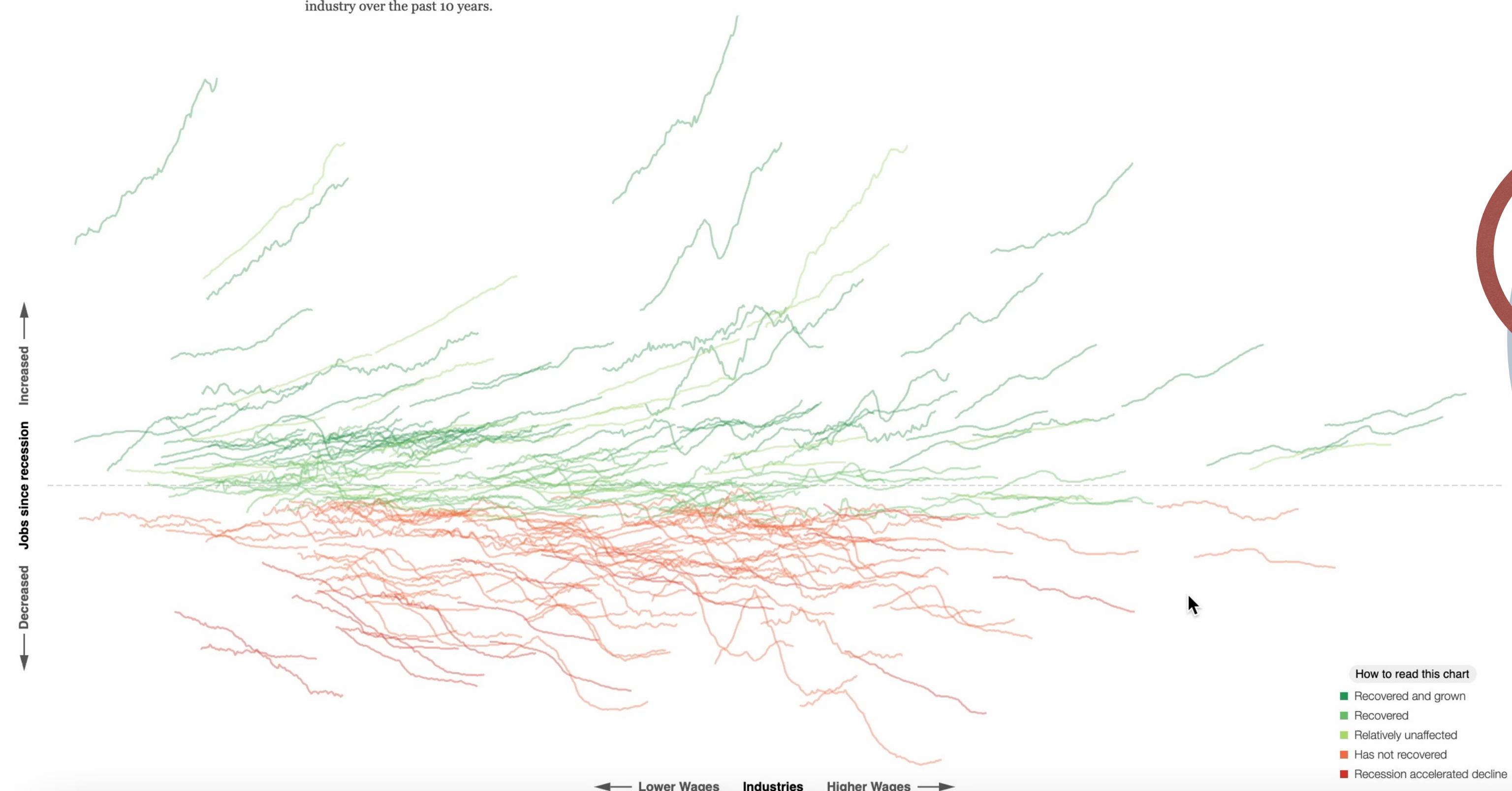
Five years since the end of the recent Recession, the economy has finally regained the nine million jobs it lost. Each line below shows how the number of jobs has changed for a particular industry over the past 10 years.



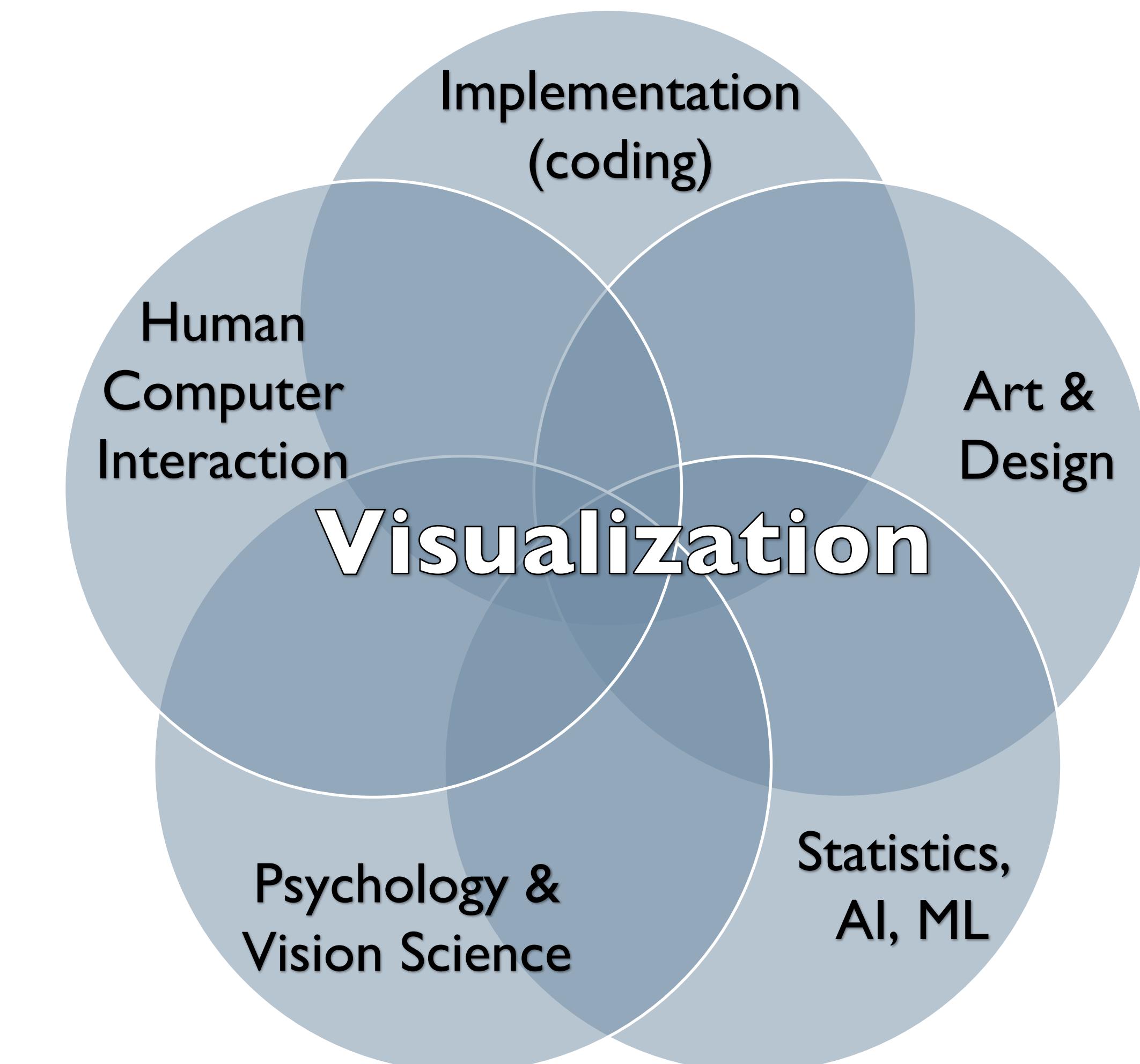
Course Overview

Analyzing trends in the economy

Five years since the end of the recent Recession, the economy has finally regained the nine million jobs it lost. Each line below shows how the number of jobs has changed for a particular industry over the past 10 years.



Course Overview



Course Structure

Classroom Technologies

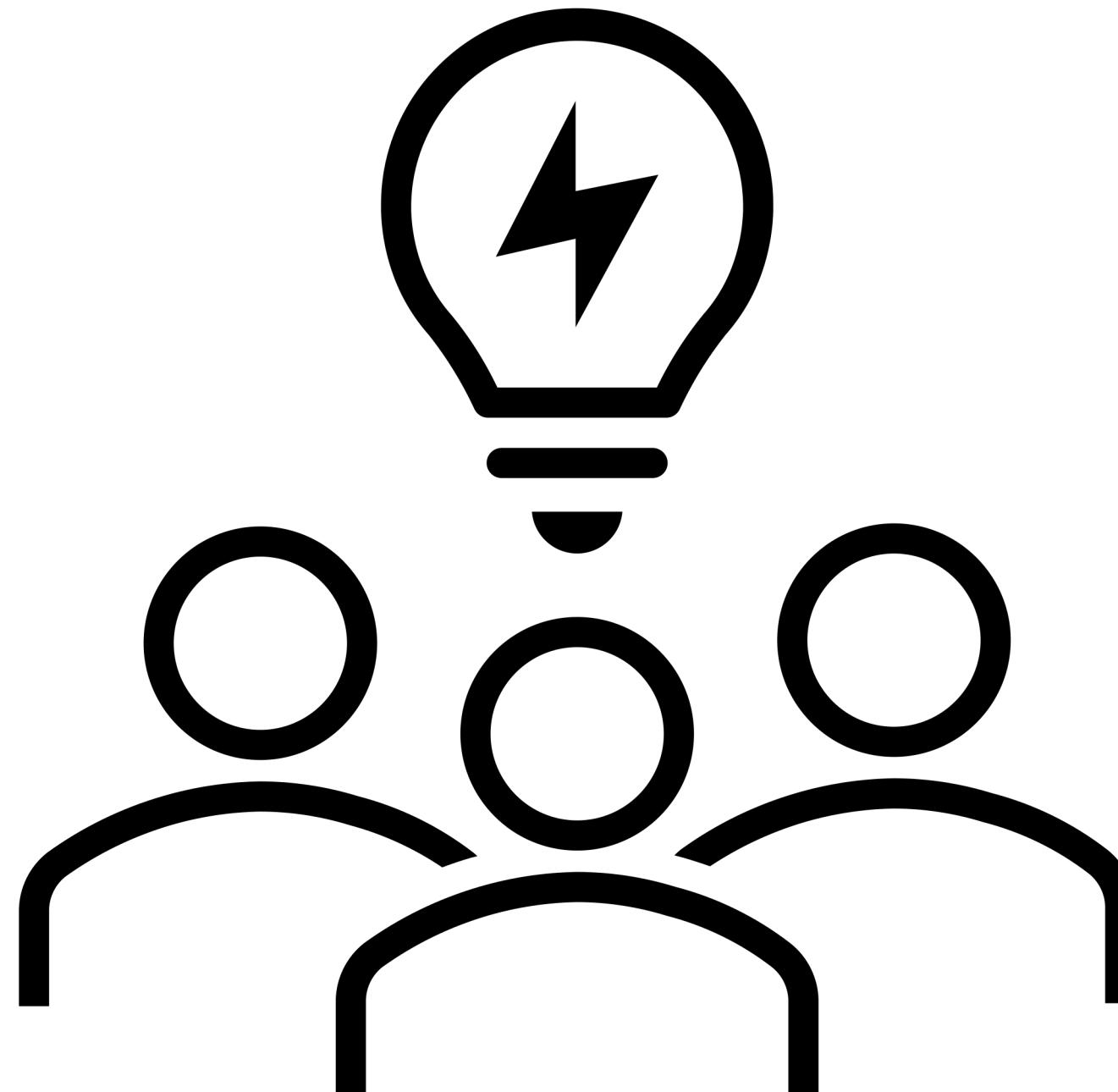
- Course Website - <https://amosca01.github.io/DS4200-F22/>
- Piazza - <https://piazza.com/class/l6o0hqf4nhx59y>
- Gradescope - <https://www.gradescope.com/courses/334321> (entry code 4VVNBG)
- GitHub Classroom – links provided in assignments

Class Structure

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

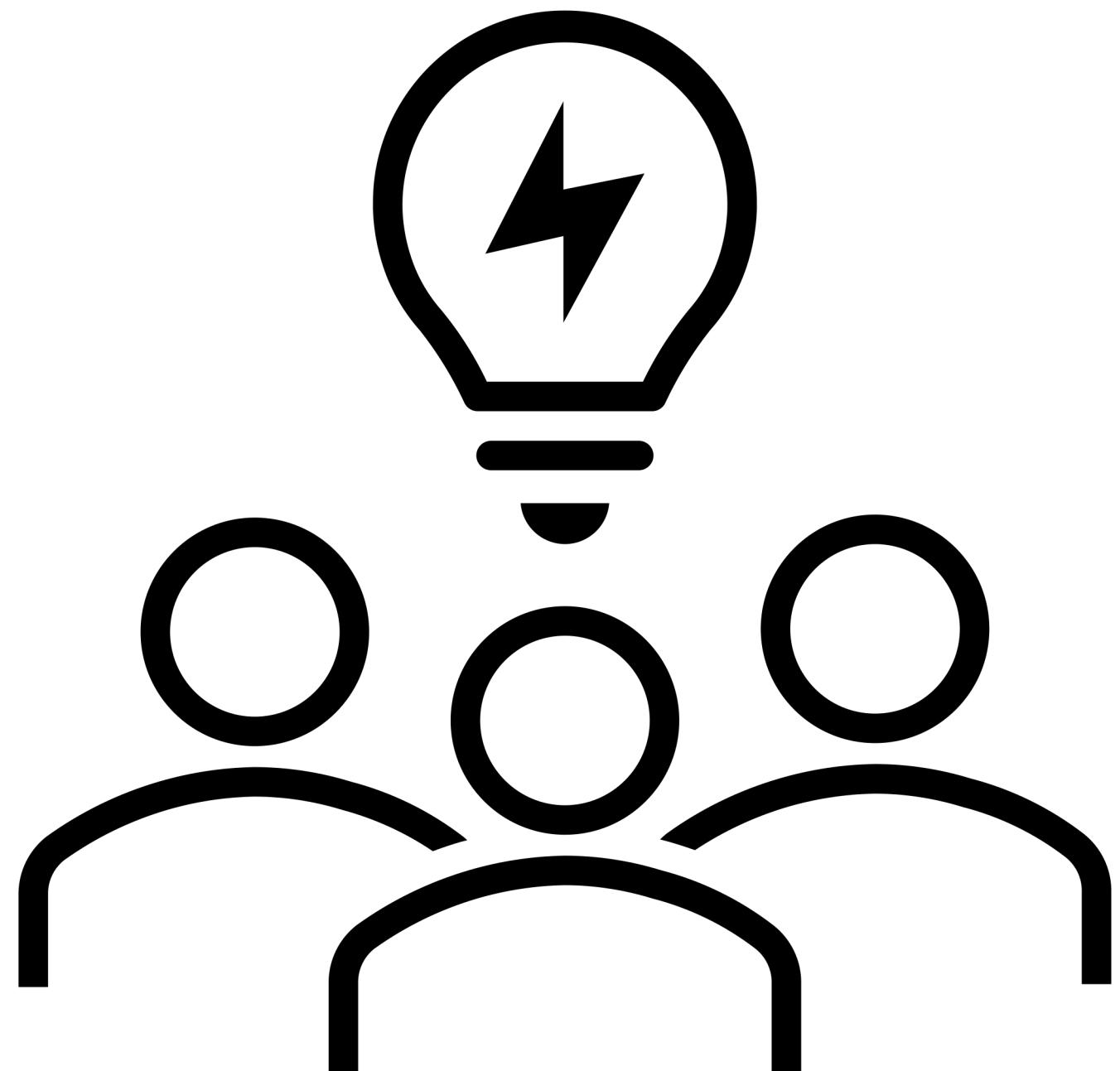
What is Visualization?

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

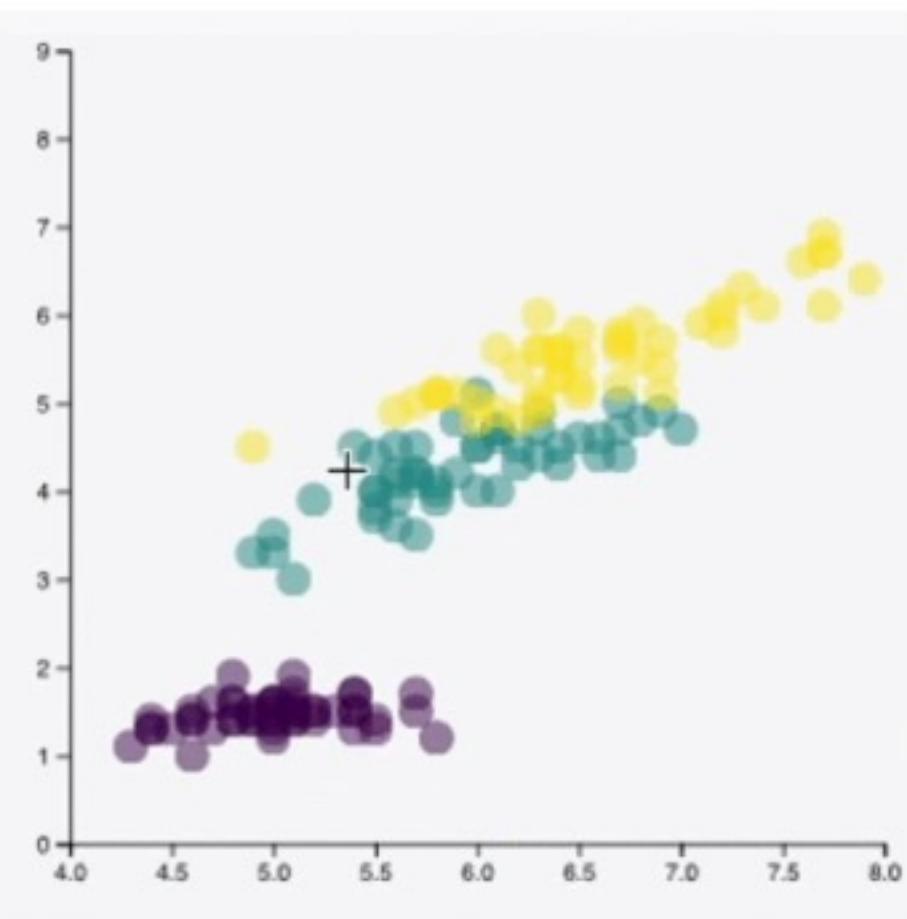
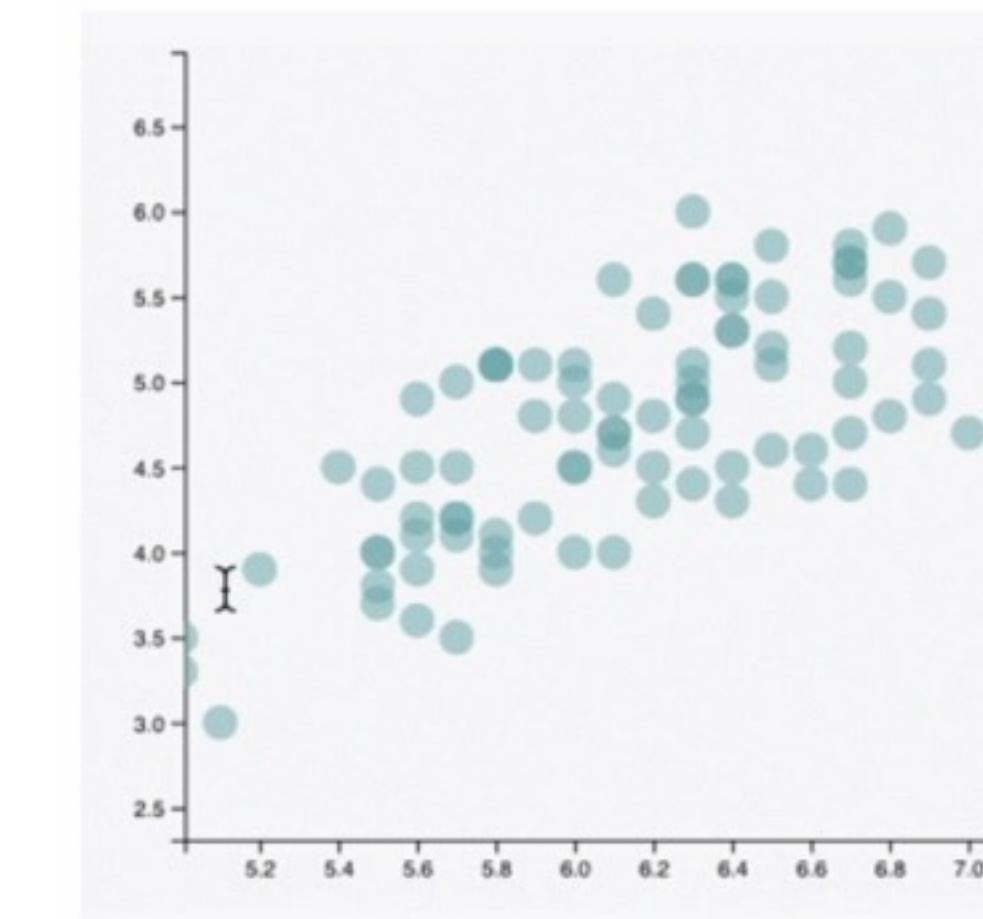
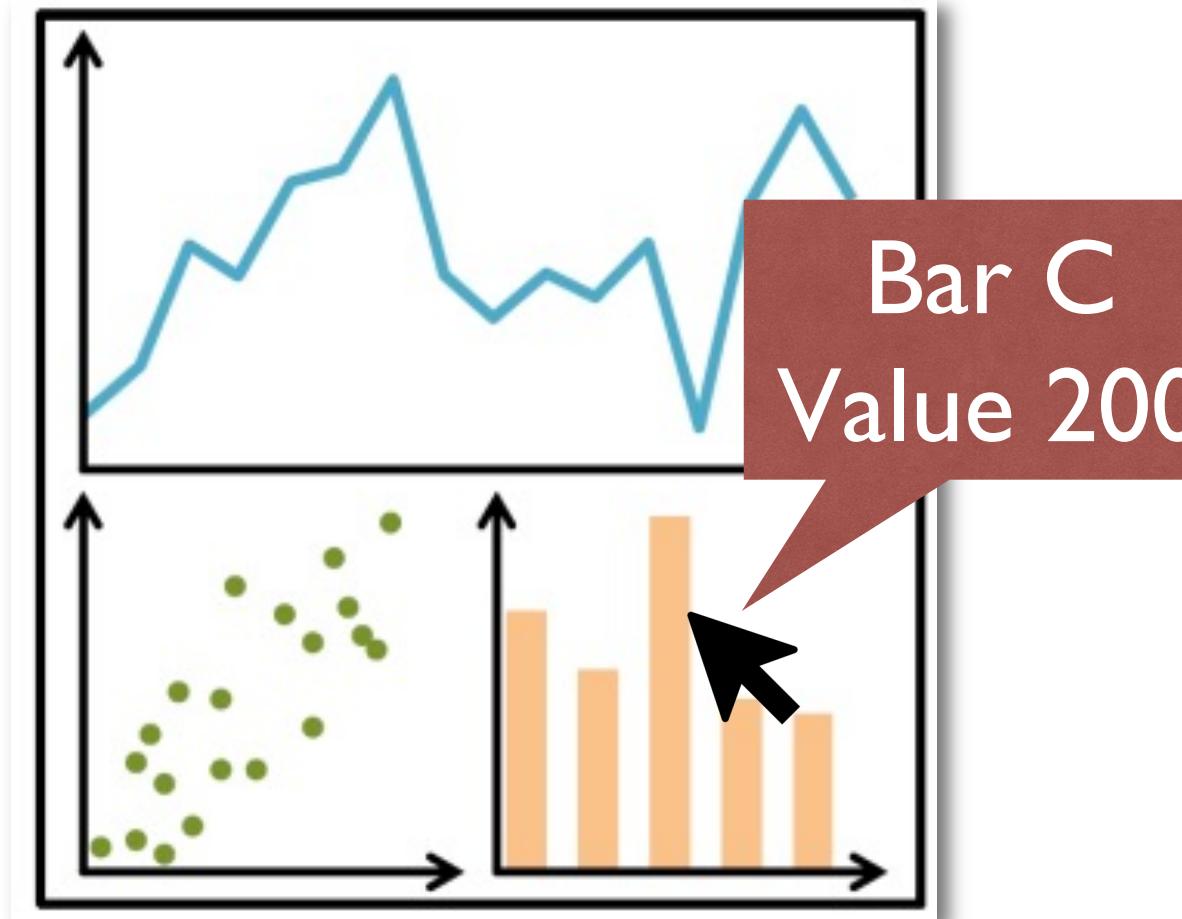
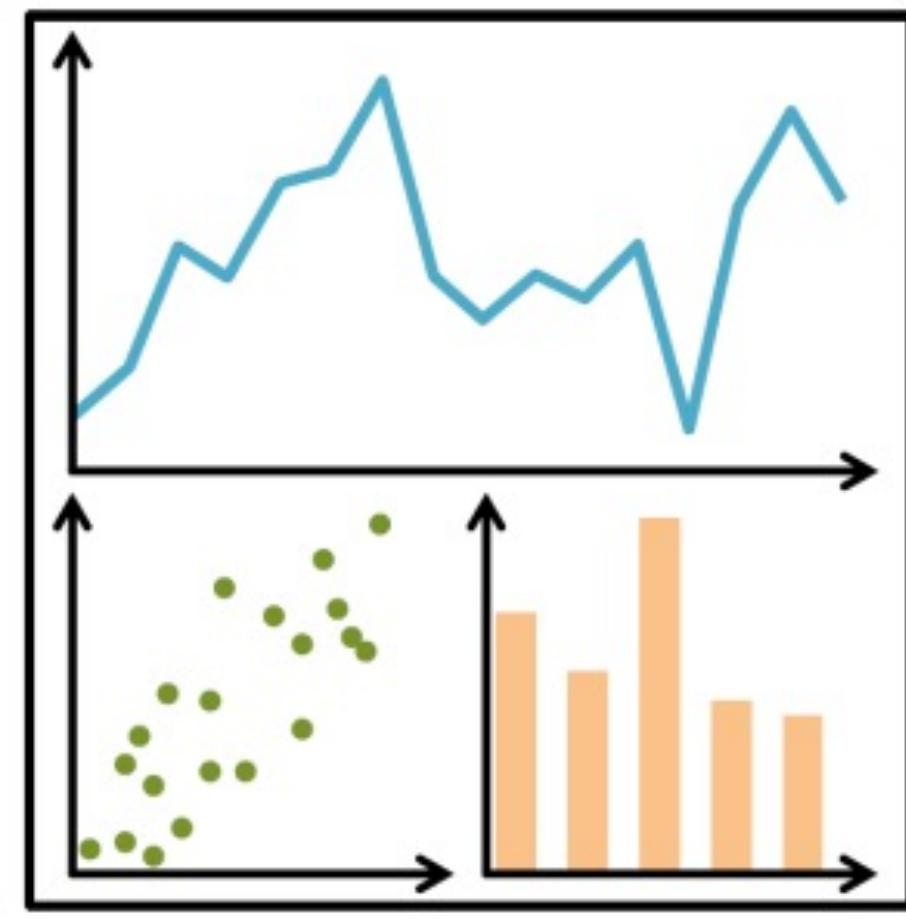
Visualization: the visual representation of data
to reinforce human cognition

Official Definition

(static or interactive)



Visualization: the visual representation of data
to reinforce human cognition

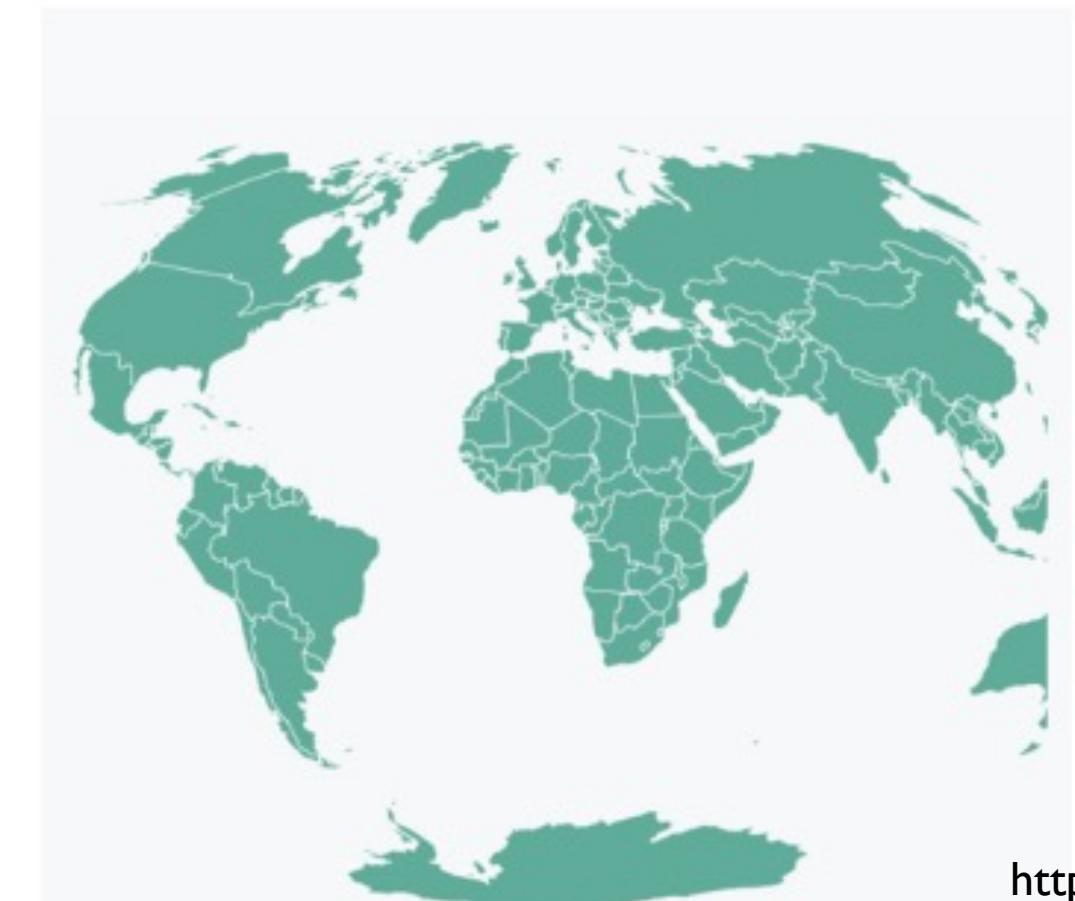
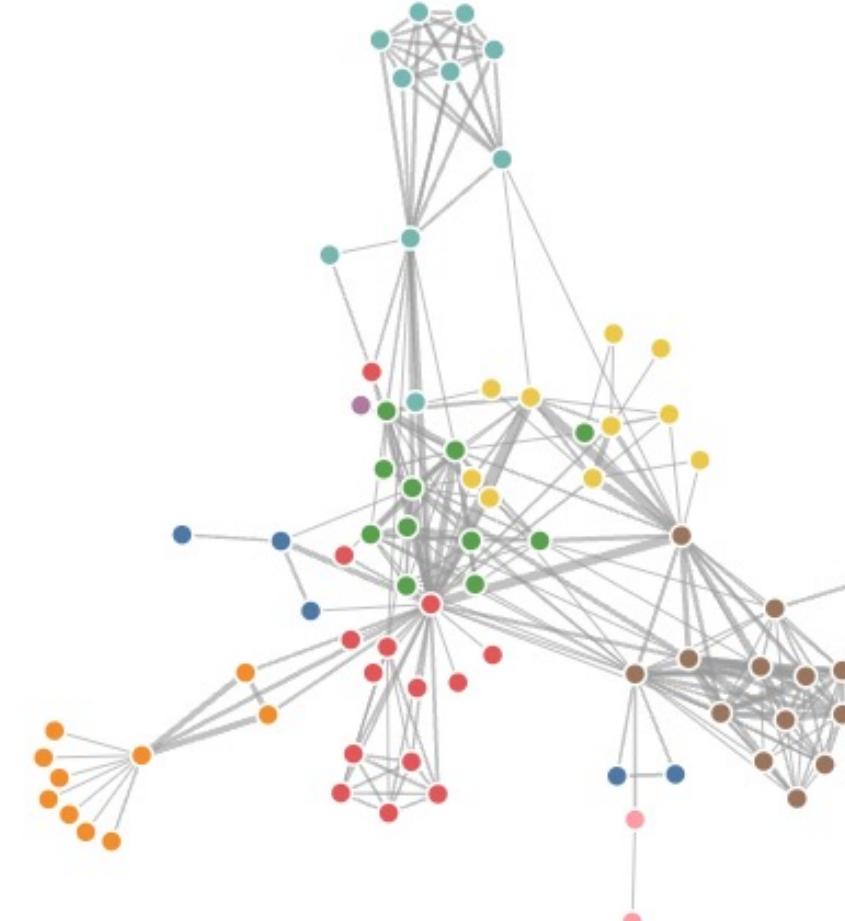
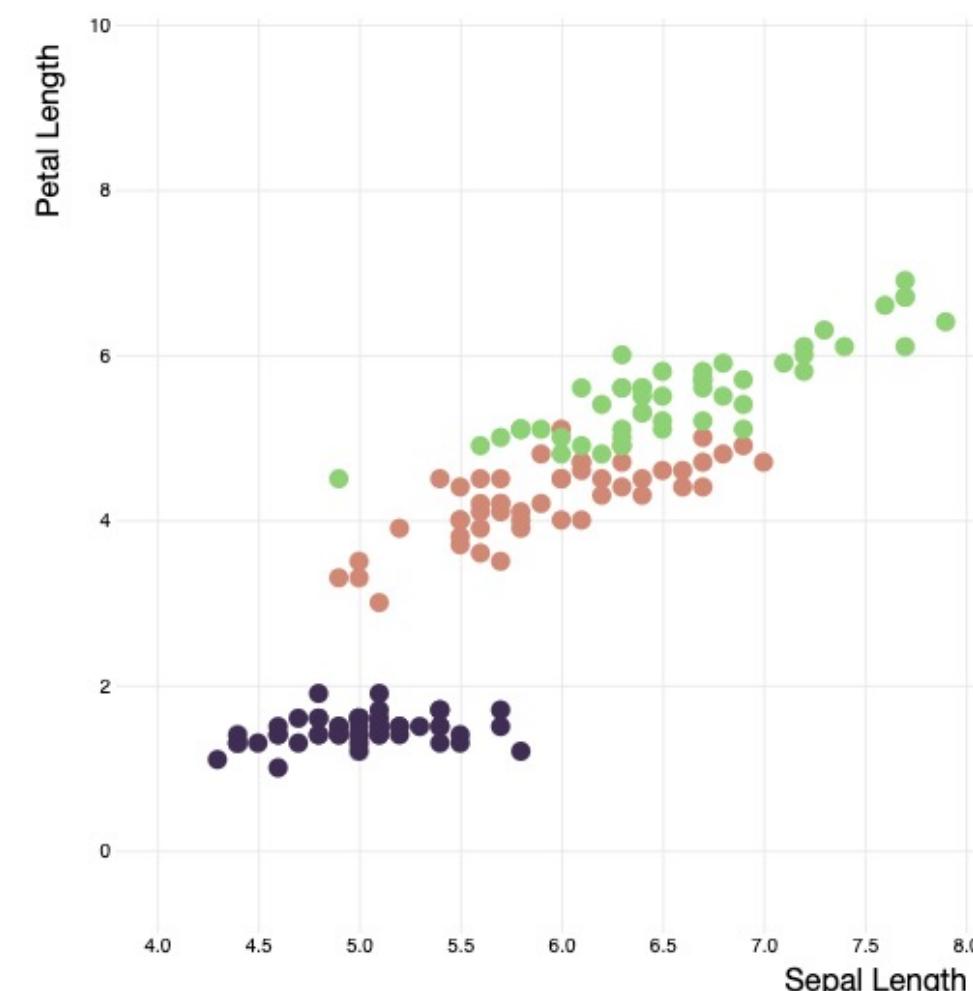


Official Definition

(static or interactive)

Visualization: the visual representation of data
to reinforce human cognition

(abstract or spatial)



Official Definition

(static or interactive)

(abstract or spatial)

Visualization: the visual representation of data
to reinforce human cognition



Why Visualization?

Why visualize data?

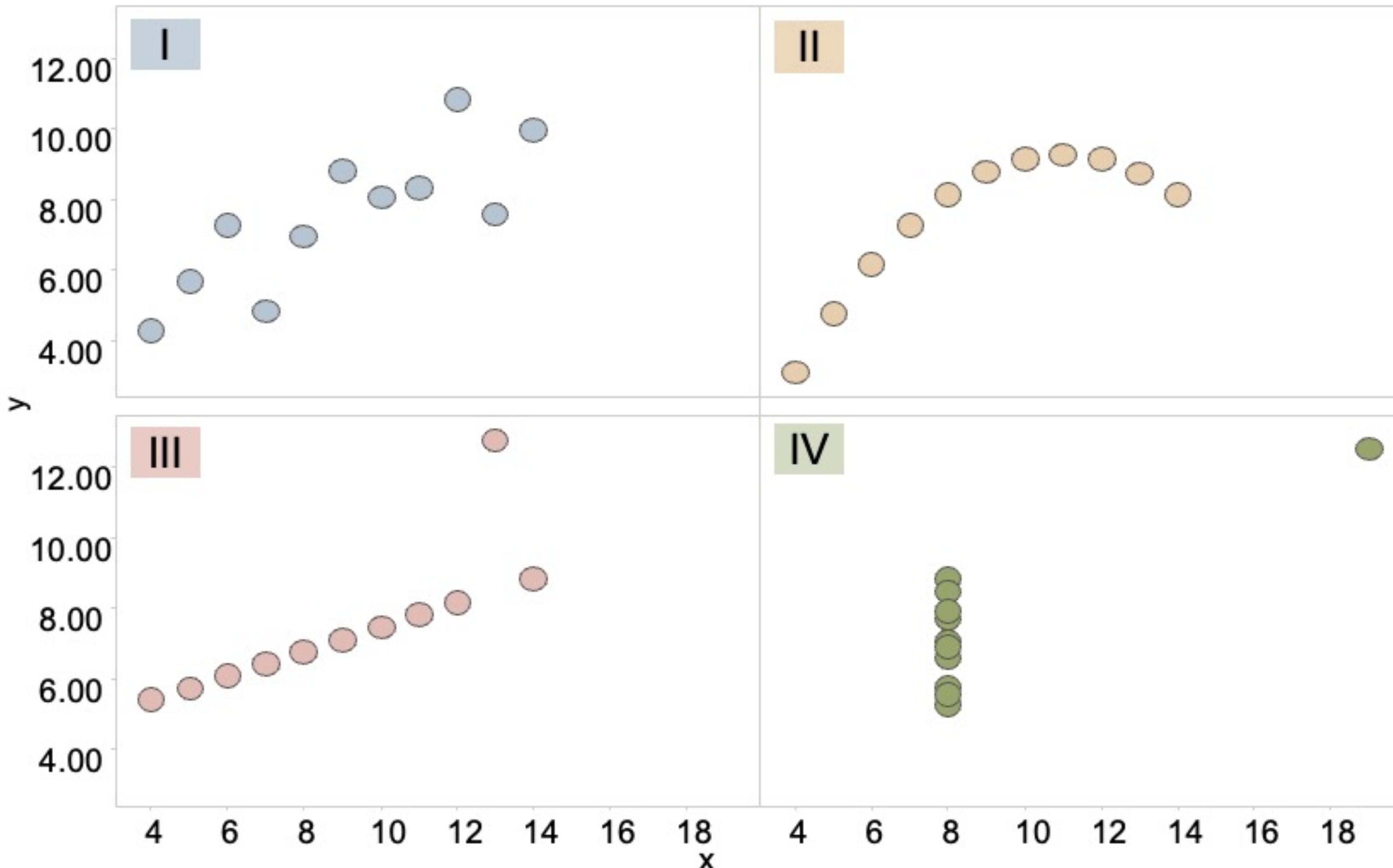
I		II		III		IV	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	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

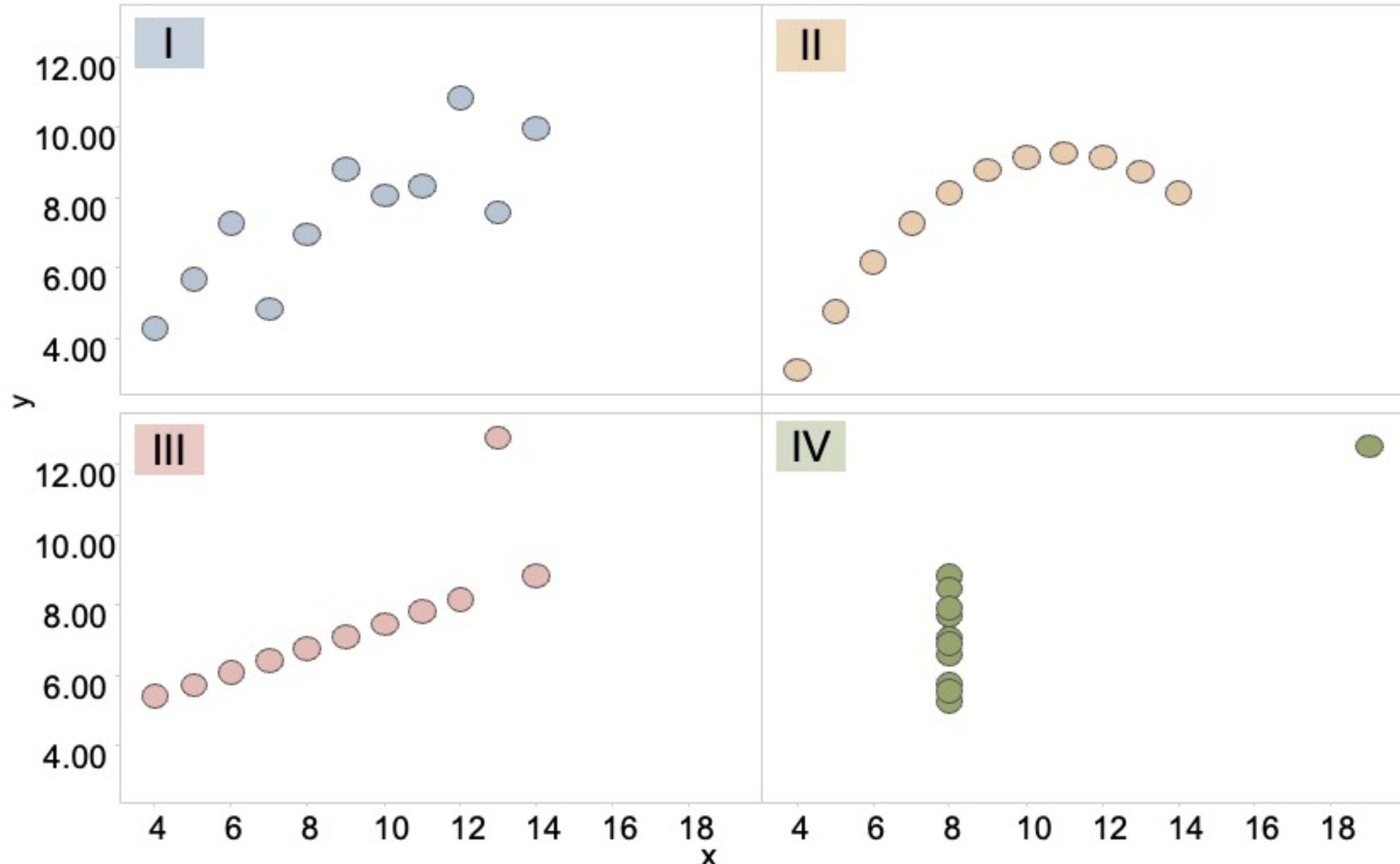
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

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?

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?

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

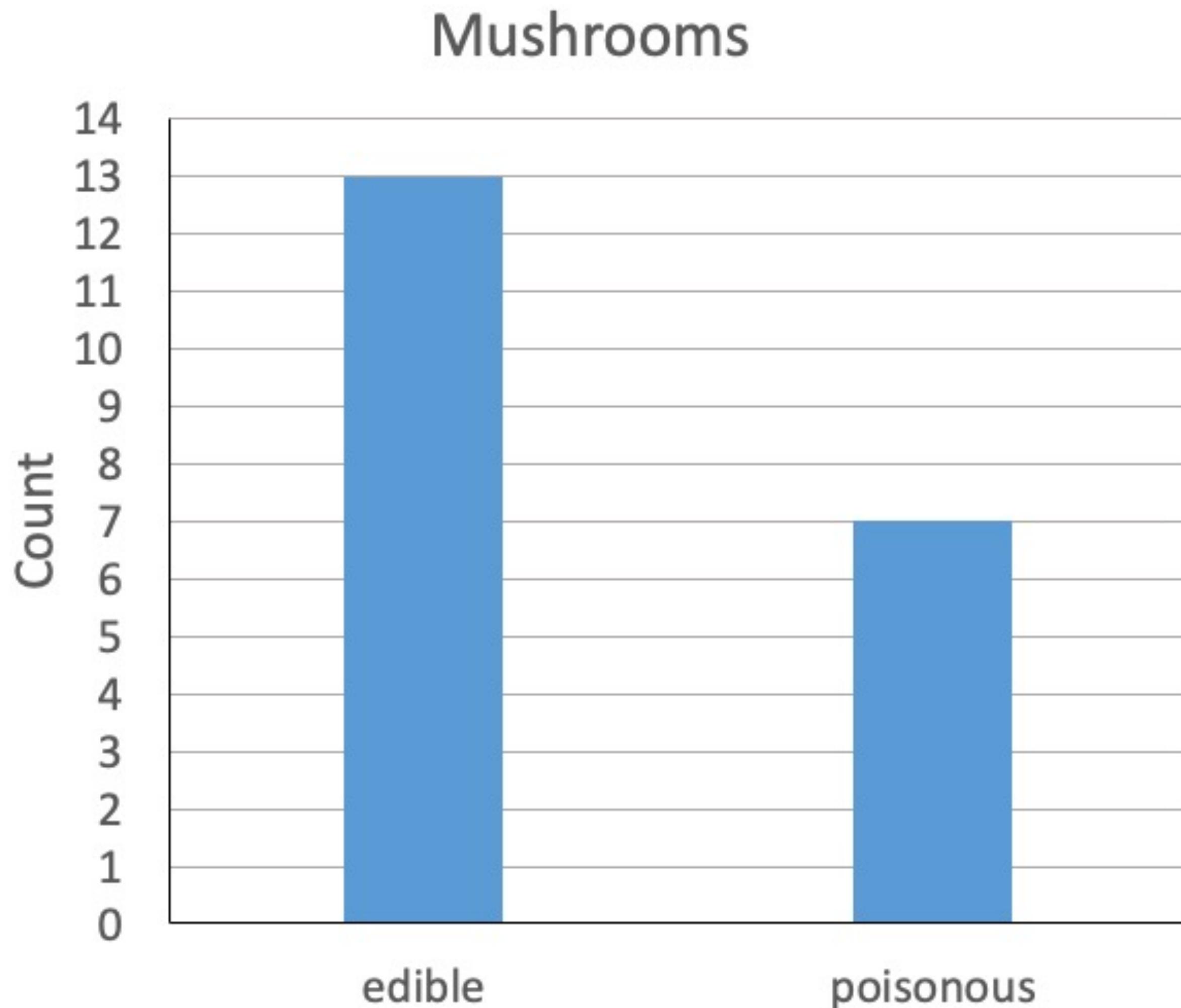
Why visualize data?

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

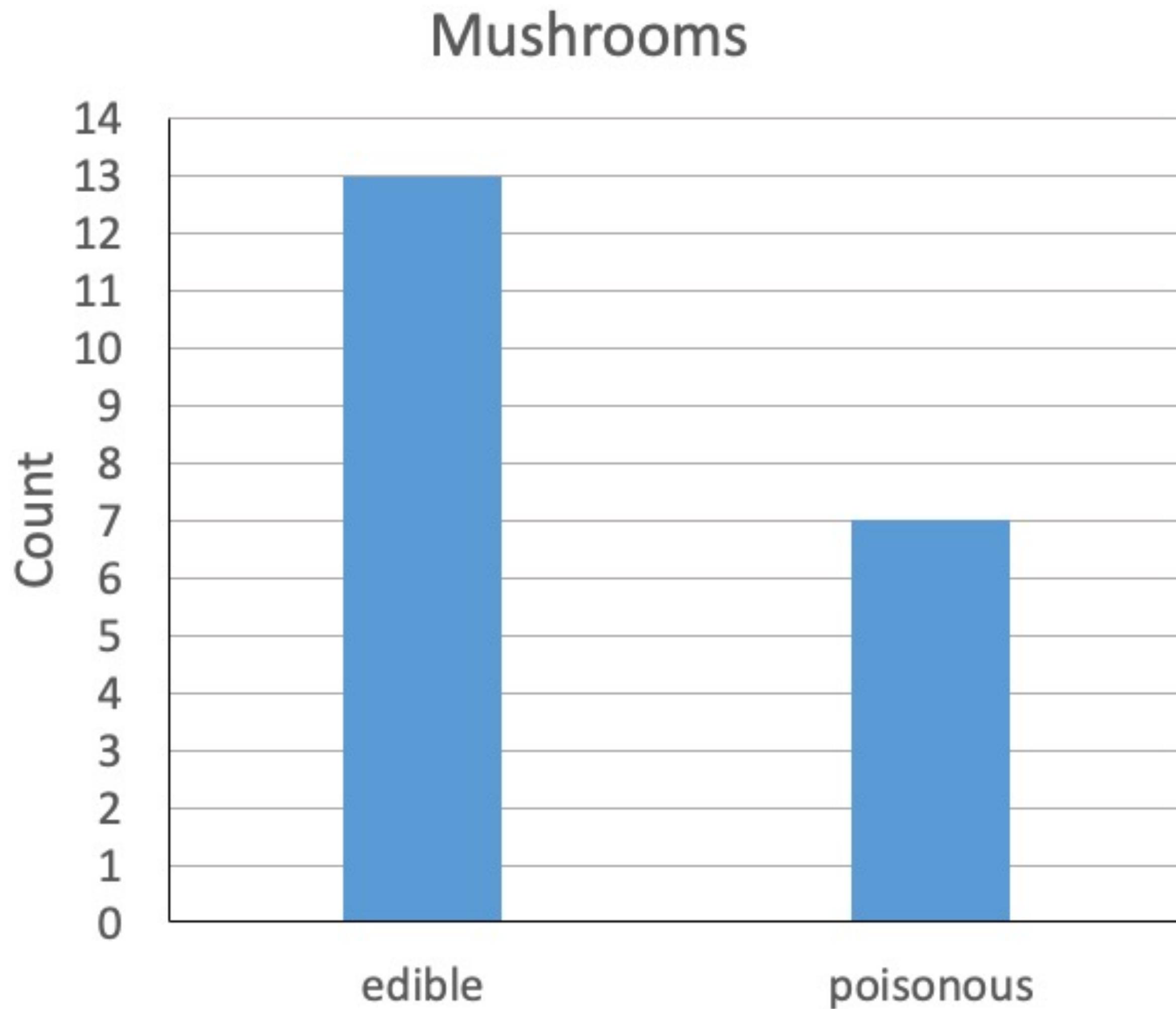
Why visualize data?

How many mushrooms are poisonous?

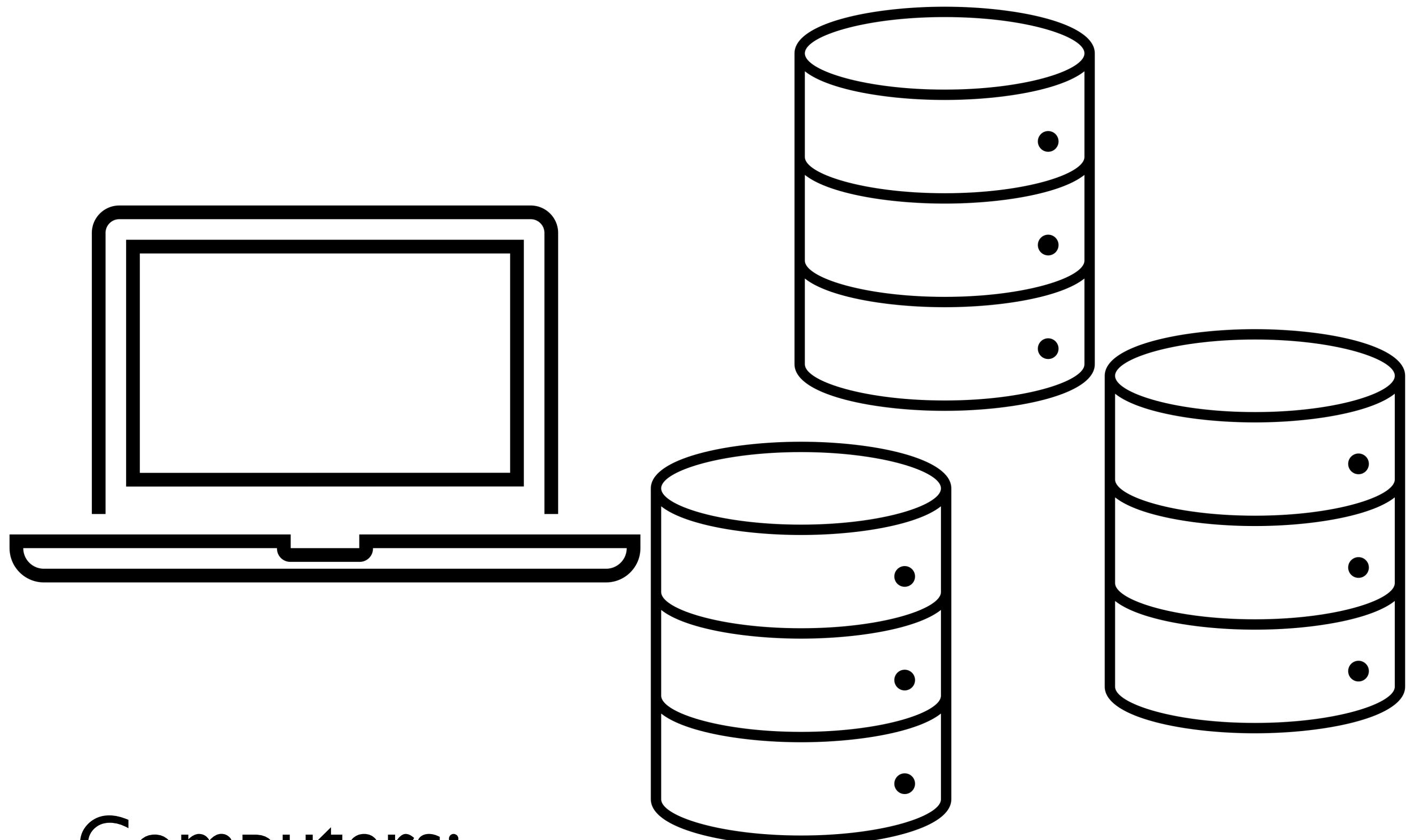
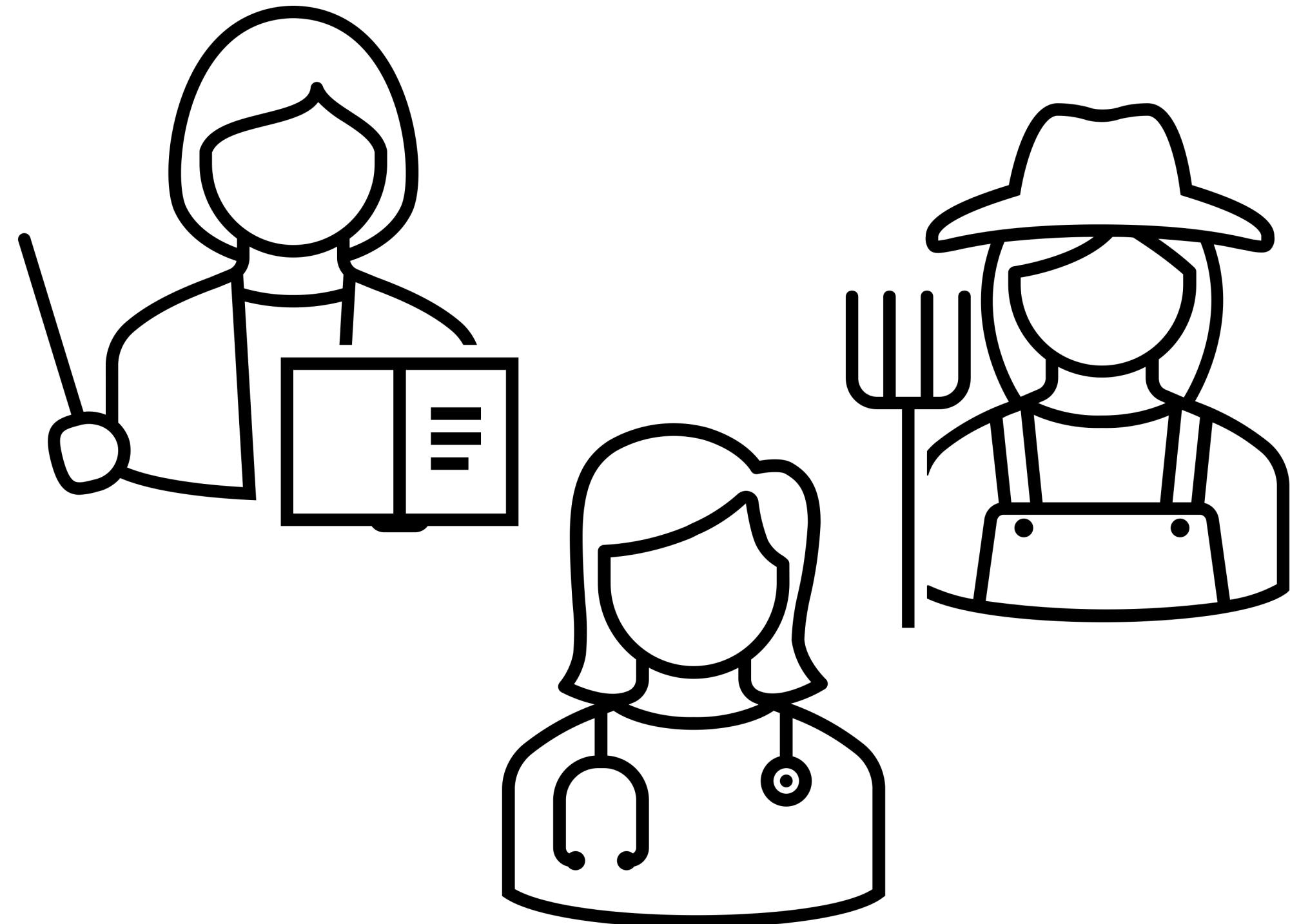


Why visualize data?

Reduce Cognitive Load



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?

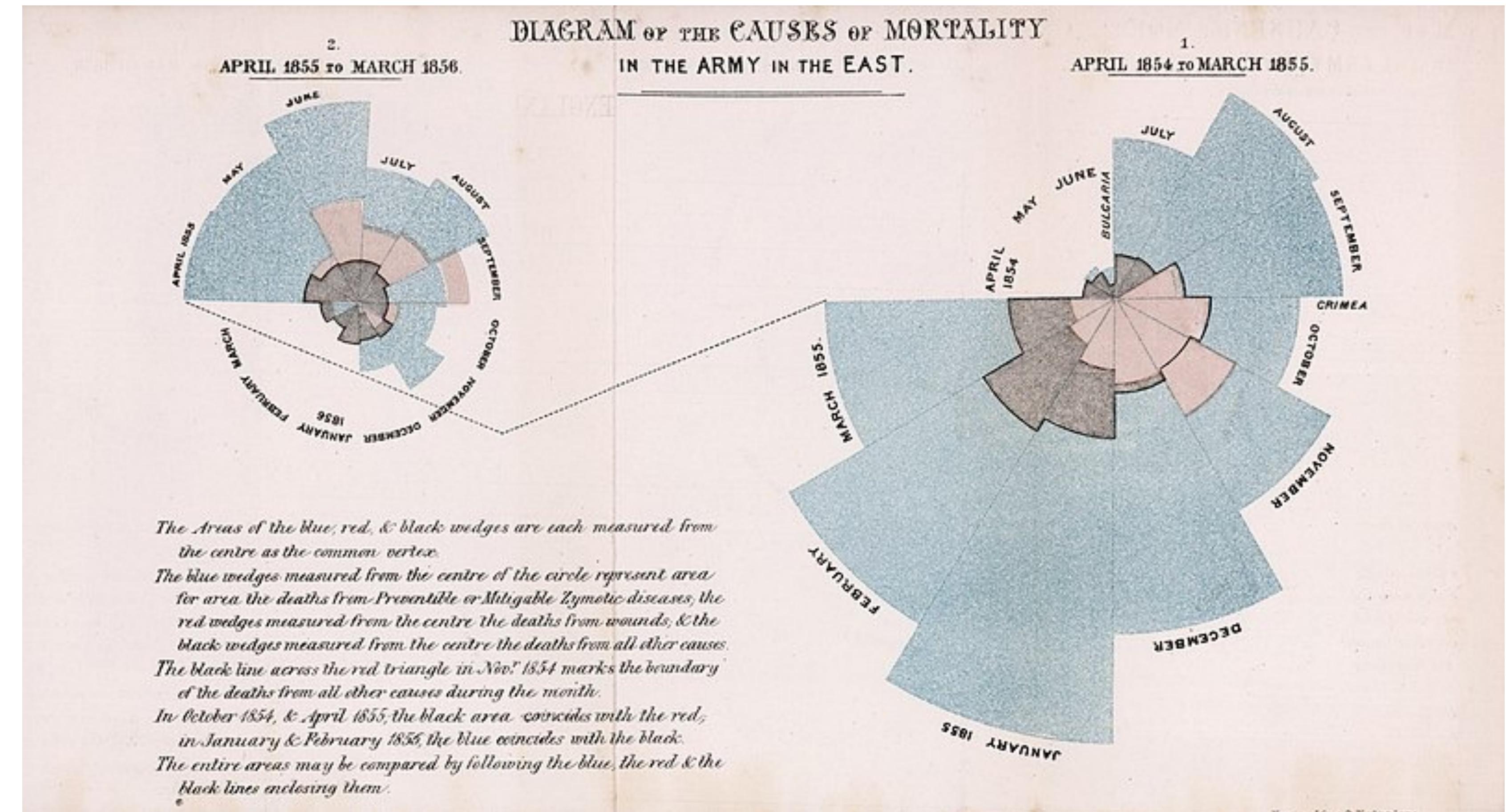
Does it really work?

John Snow's Cholera Map



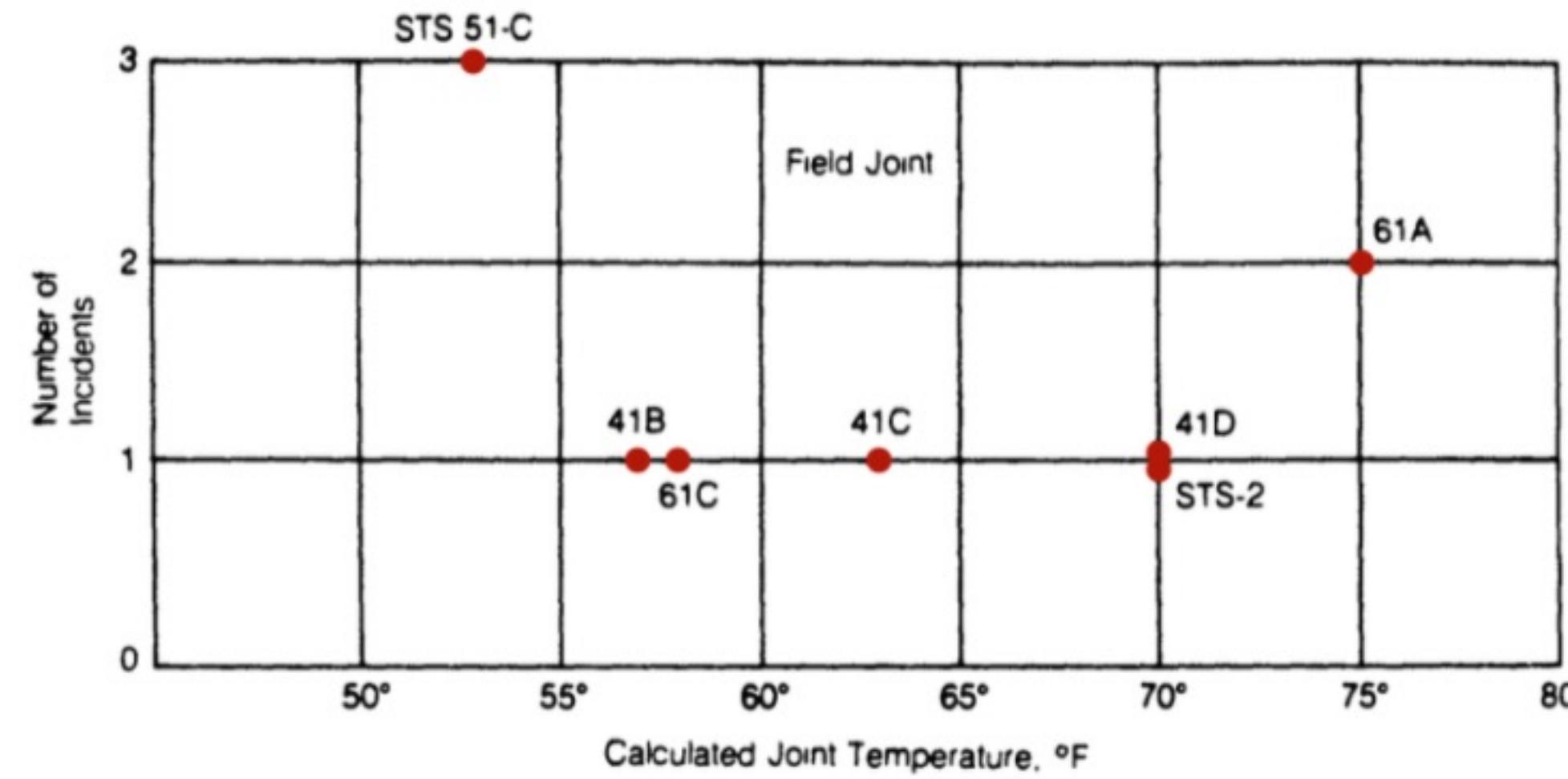
Does it really work?

Florence Nightingale's Rose



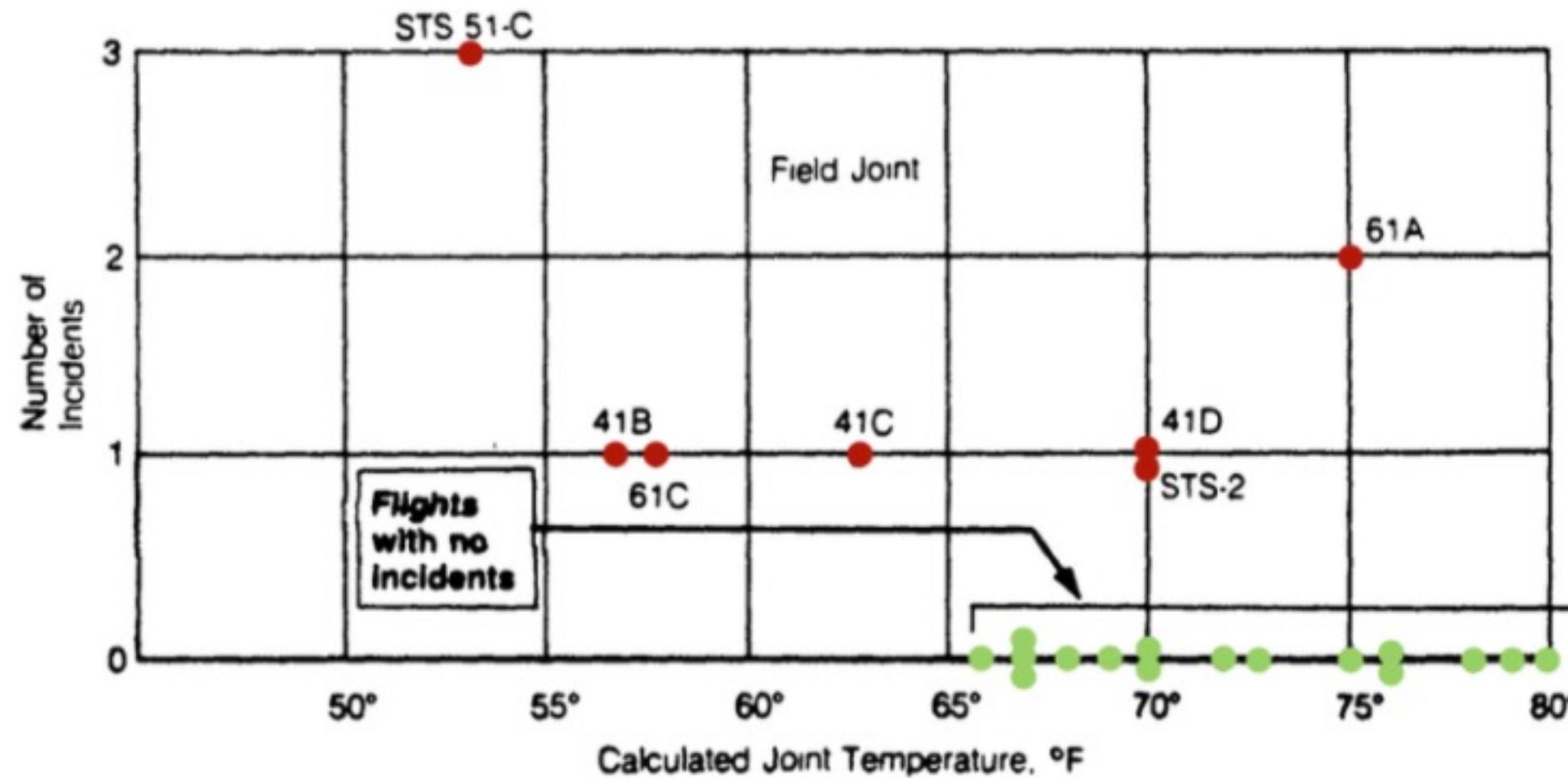
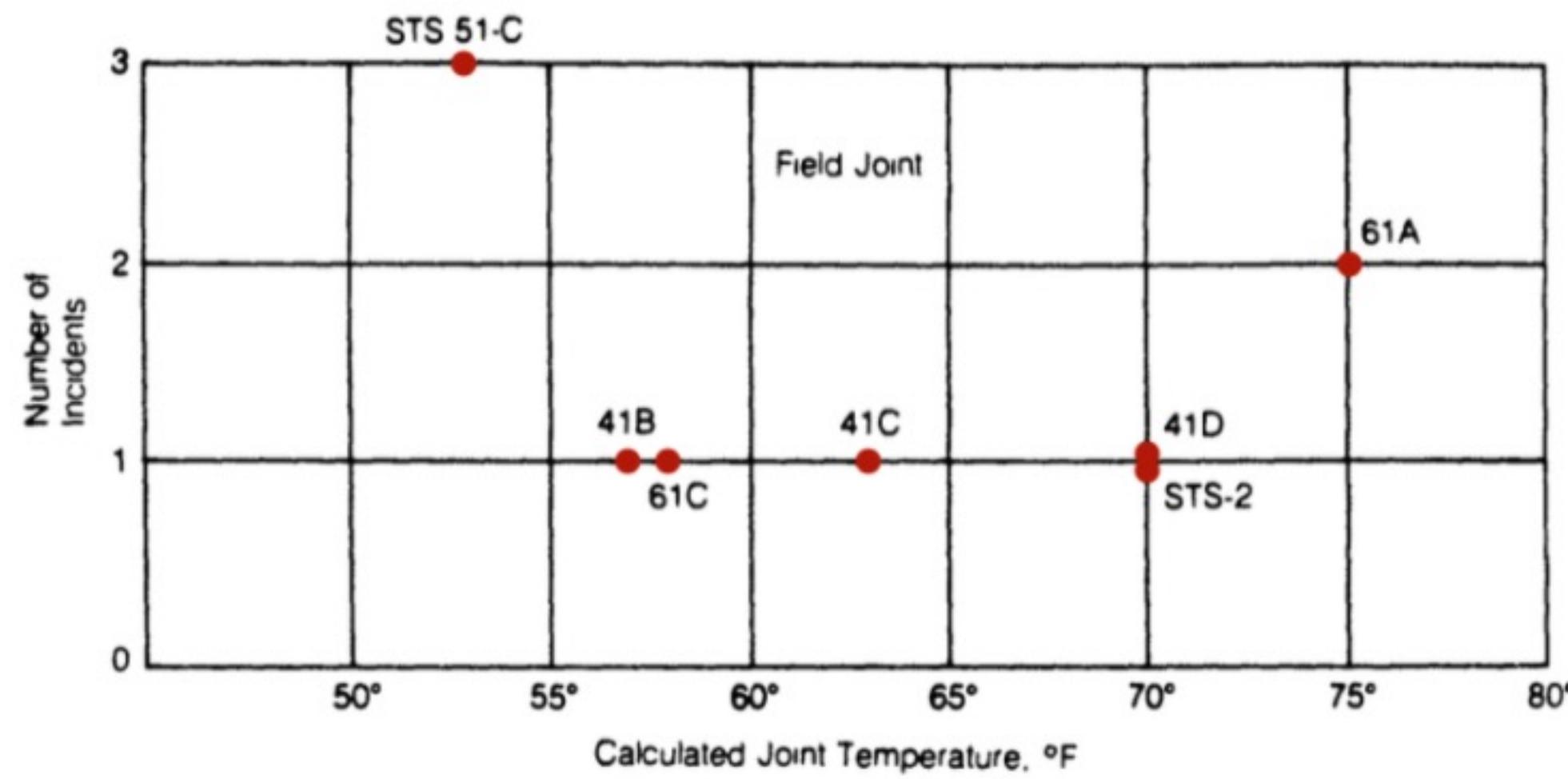
Does it really work?

Challenger Shuttle



Does it really work?

Challenger Shuttle



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

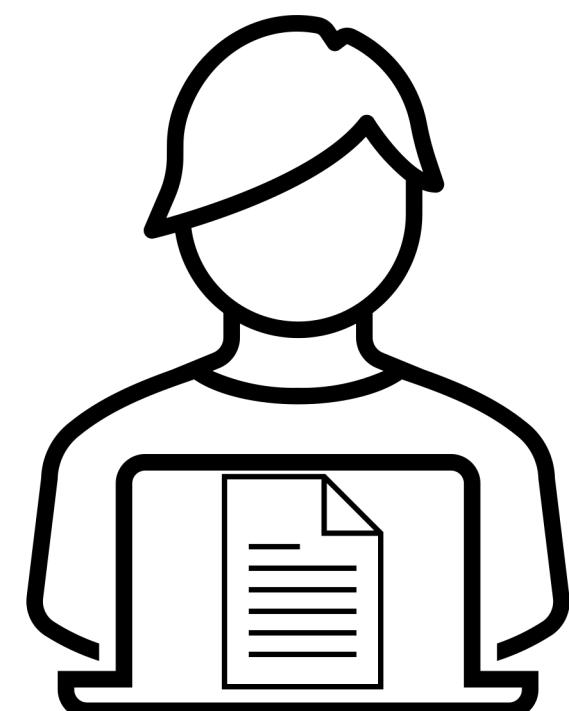
stores content and tracks changes

content stored in multiple places

Introduction to Git



Remote = copy on the server



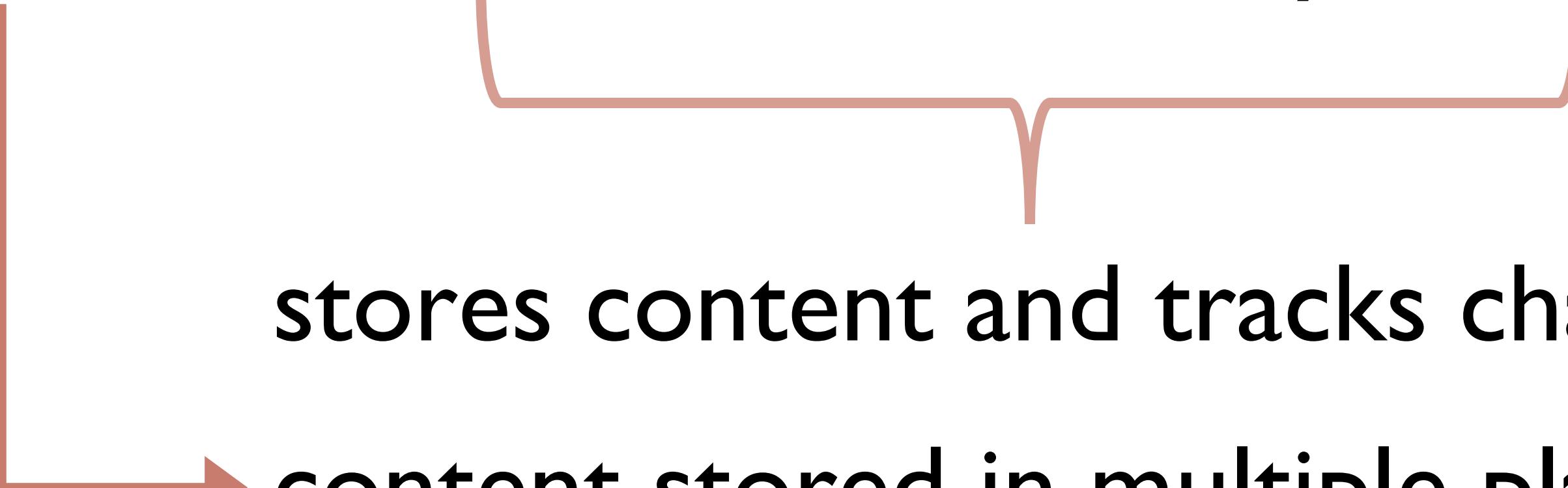
Local = copy on your computer

What is Git?

A Distributed Version Control System

stores content and tracks changes

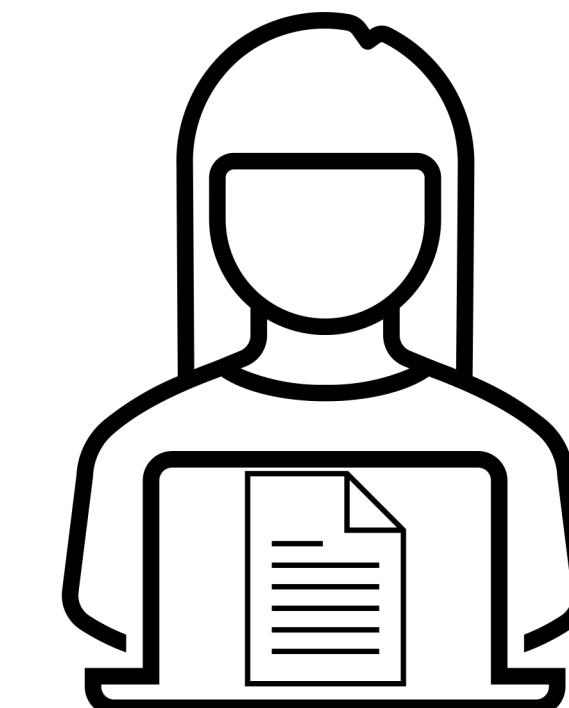
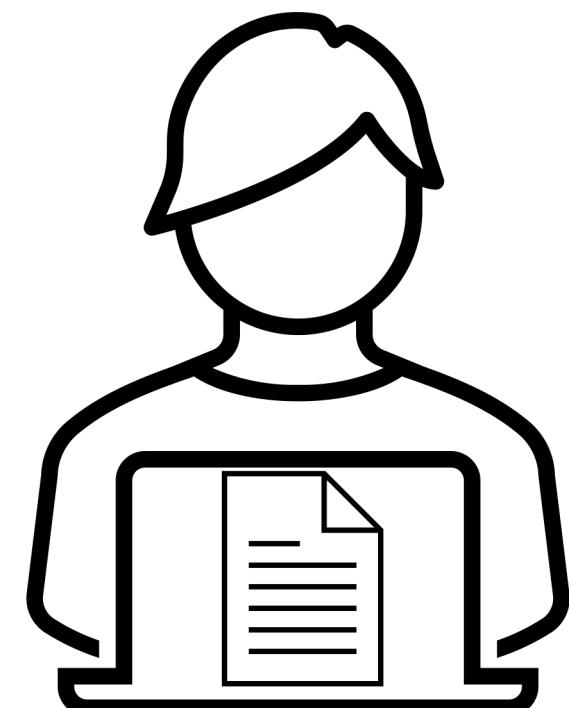
content stored in multiple places



Introduction to Git



Remote = copy on the server



Local = copy on your computer

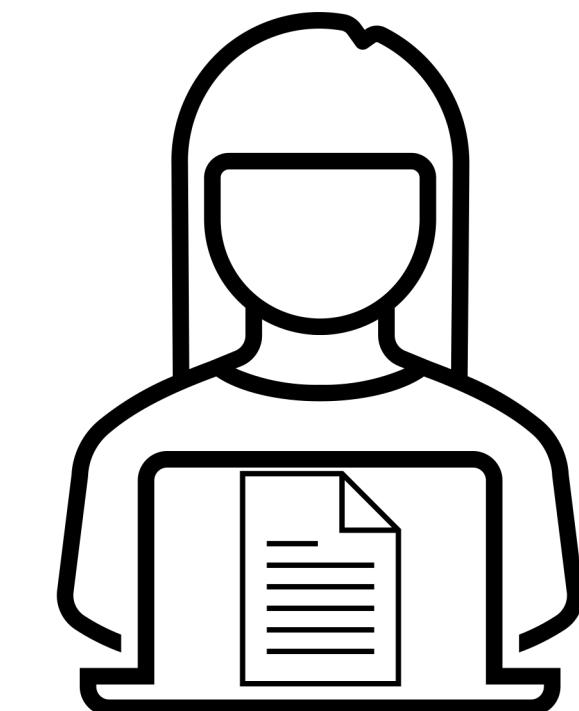
Why use Git?

- 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

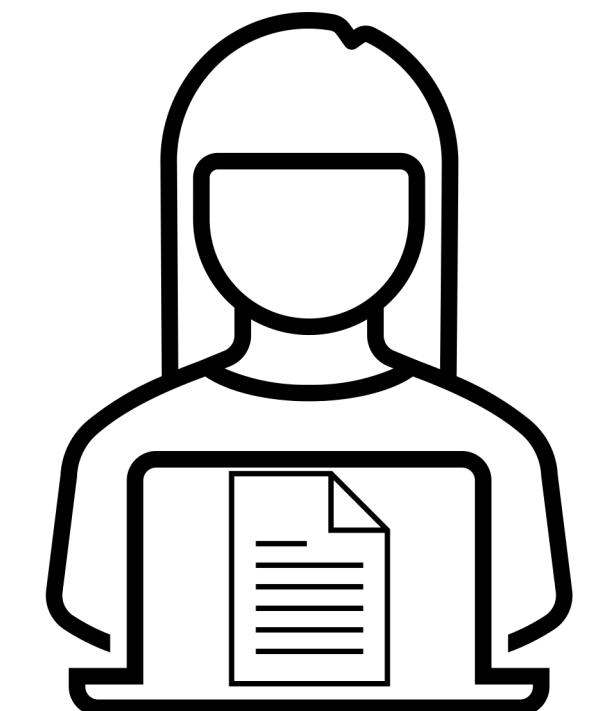
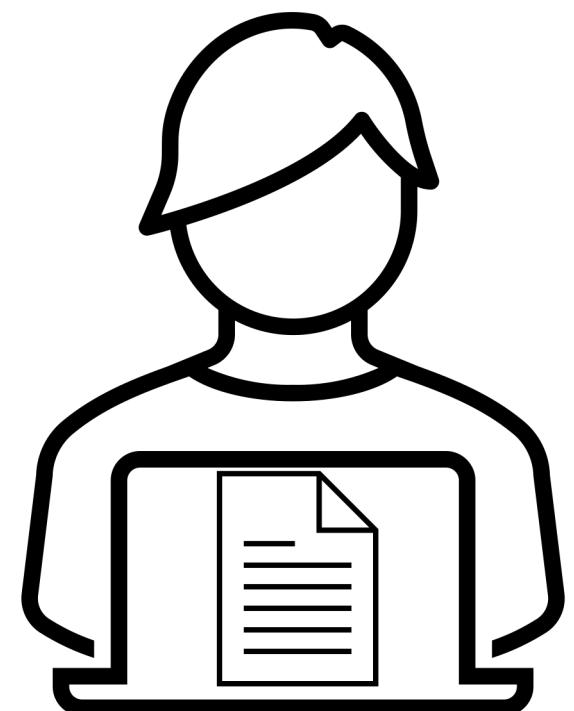
Ex. Suppose you finished part 1 of your homework. You move on to part 2 but part way through realize you broke part 1 and cannot figure out how.

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?

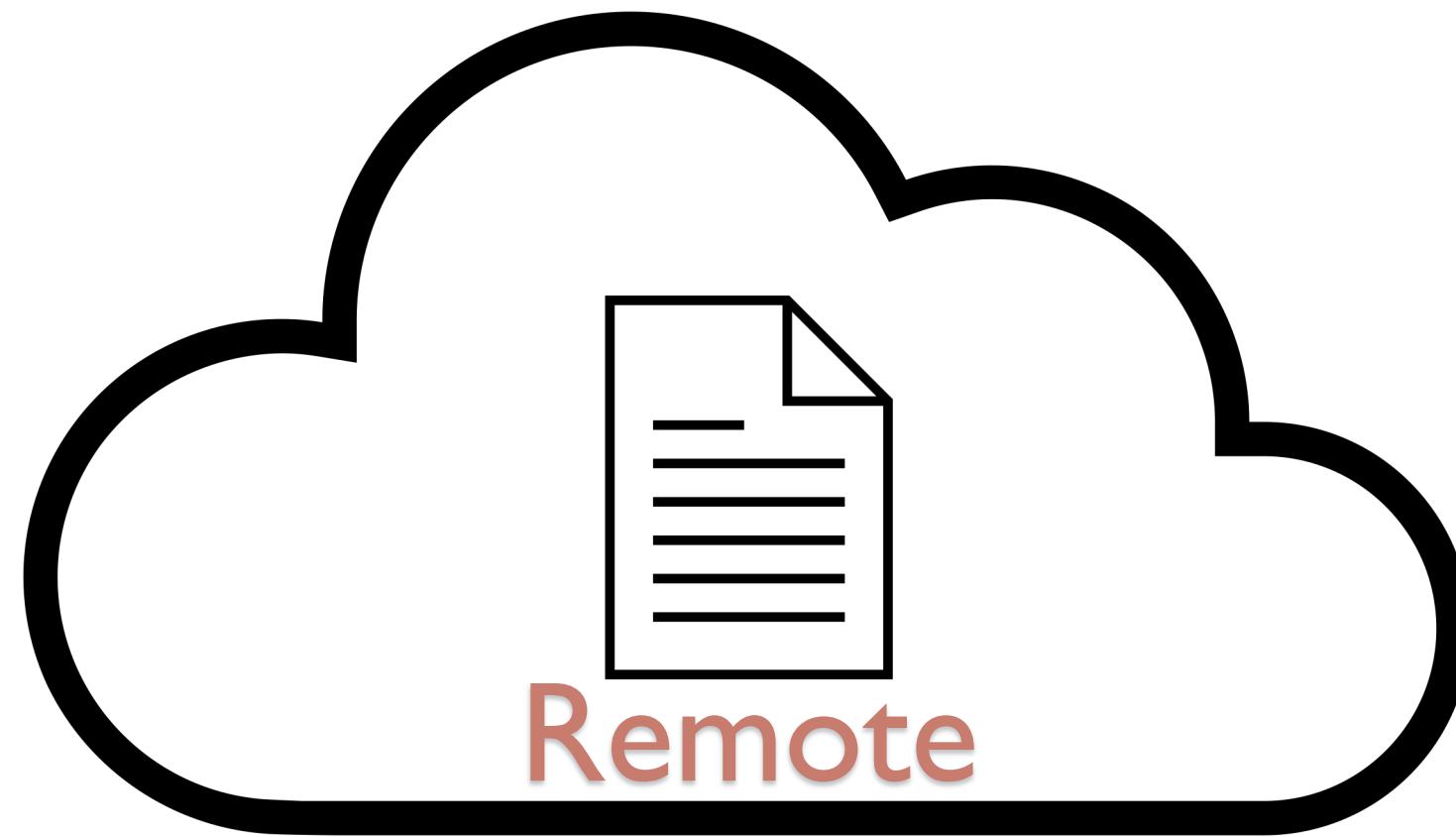
To keep backups of your work

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

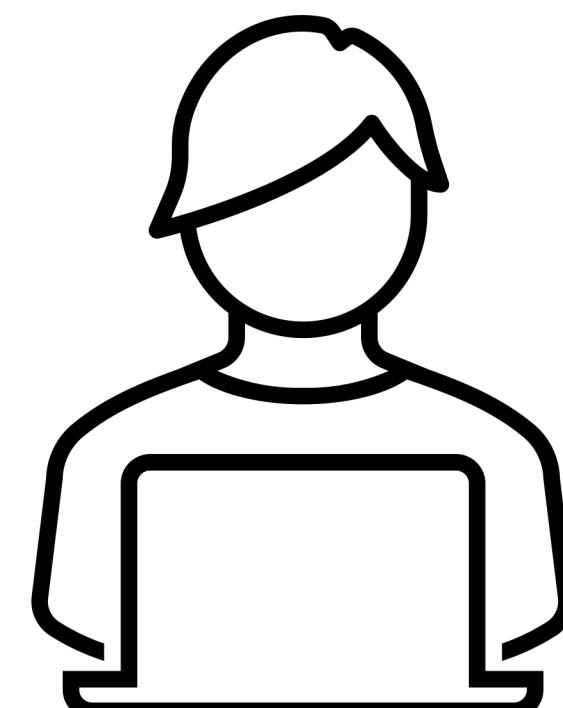


How to use Git?

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



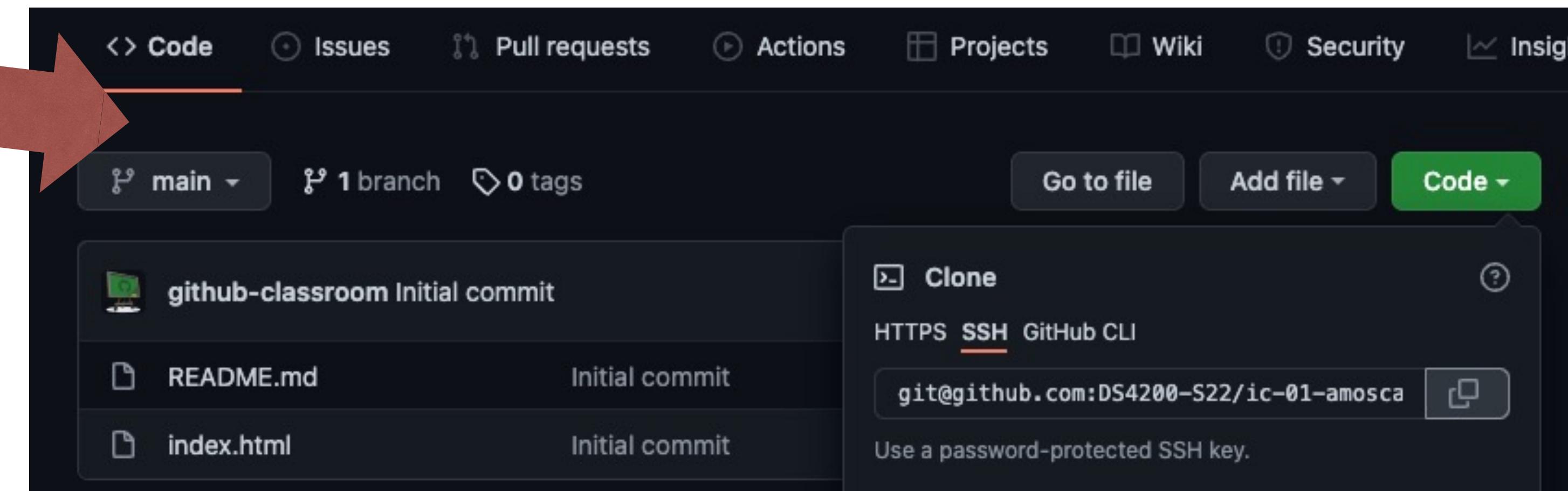
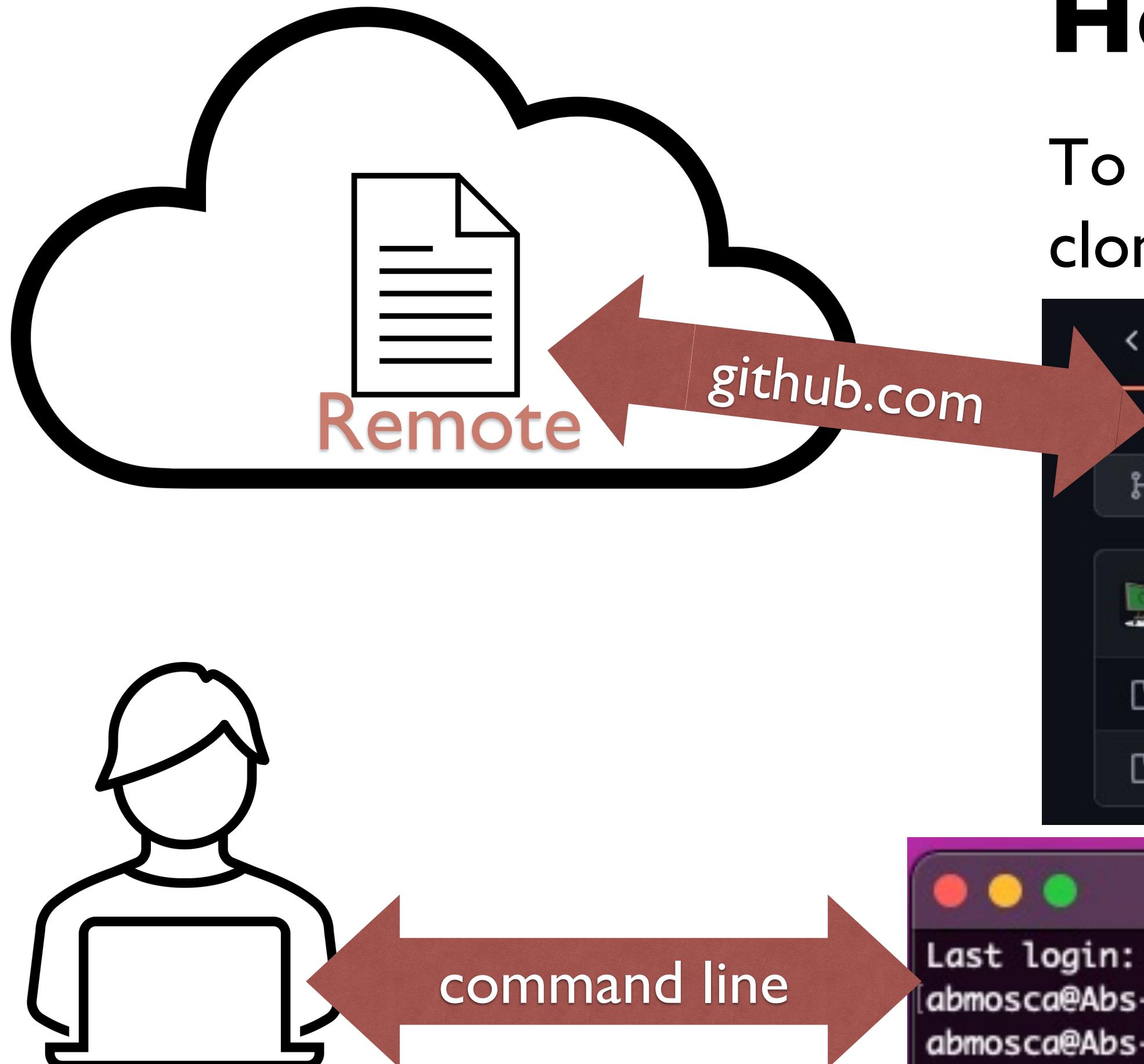
How to use Git?

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

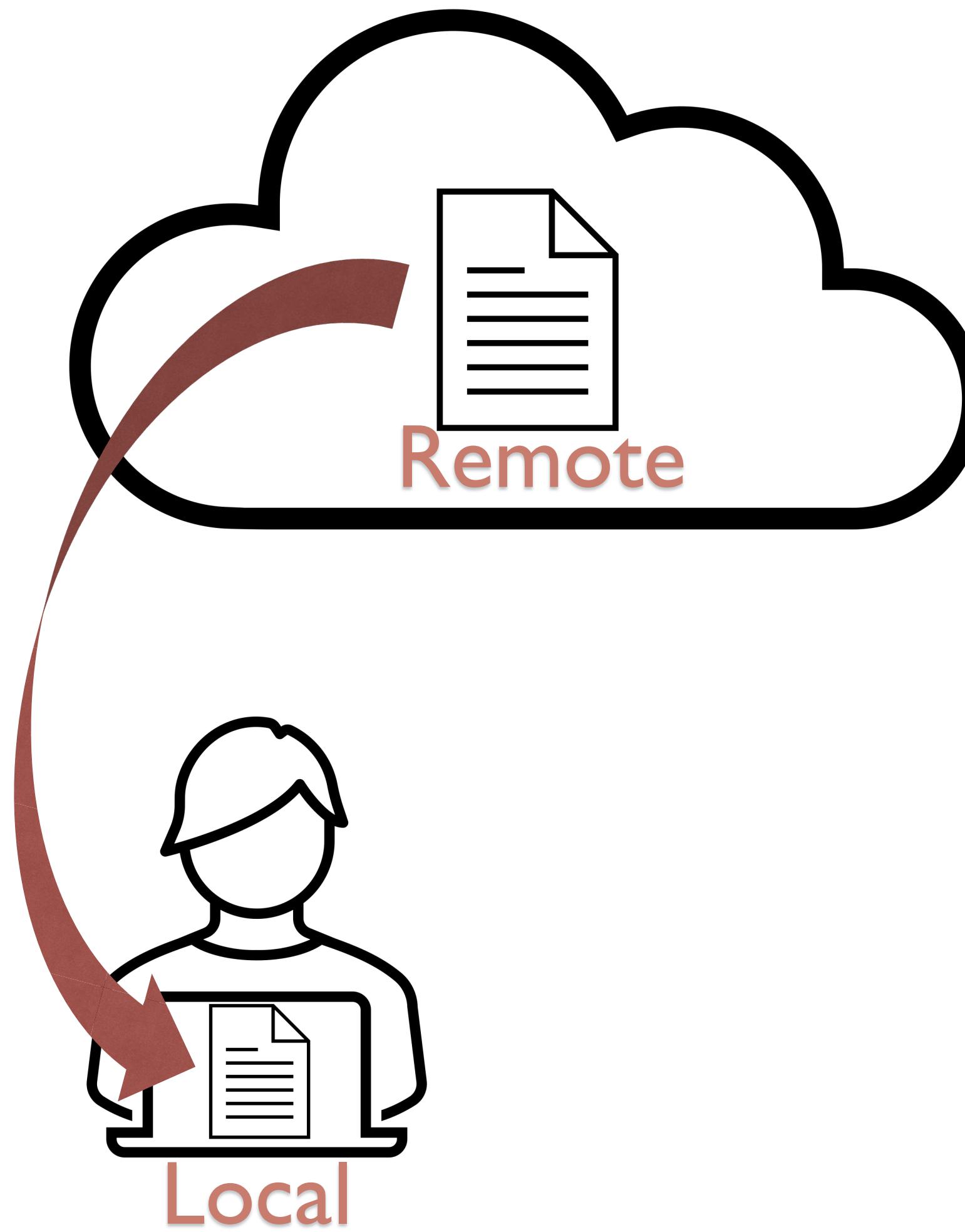
How to use Git?

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



A screenshot of a terminal window titled "DS4200-S22 — -zsh — 100x47". The window shows the user's last login information: "Last login: Wed Jan 19 08:58:32 on console". The user then runs the command "cd Documents/Documents/DS4200-S22" followed by "git clone git@github.com:DS4200-S22/ic-01-amosca01.git".

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).

The top half of the image shows a screenshot of a GitHub repository page for "github-classroom". The page displays basic statistics: 1 branch, 0 tags. Below this, there's a list of files: "github-classroom Initial commit", "README.md Initial commit", and "index.html Initial commit". On the right side of the page, a "Clone" button is highlighted, showing options for "HTTPS", "SSH", and "GitHub CLI", along with the URL "git@github.com:DS4200-S22/ic-01-amosca". A note below says "Use a password-protected SSH key.".

The bottom half shows a terminal window titled "DS4200-S22 — -zsh — 100x47". The terminal output is as follows:

```
Last login: Wed Jan 19 08:58:32 on console
[abmosca@Abs-MacBook-Pro ~ % cd Documents/Documents/DS4200-S22
abmosca@Abs-MacBook-Pro DS4200-S22 % git clone git@github.com:DS4200-S22/ic-01-amosca01.git
```

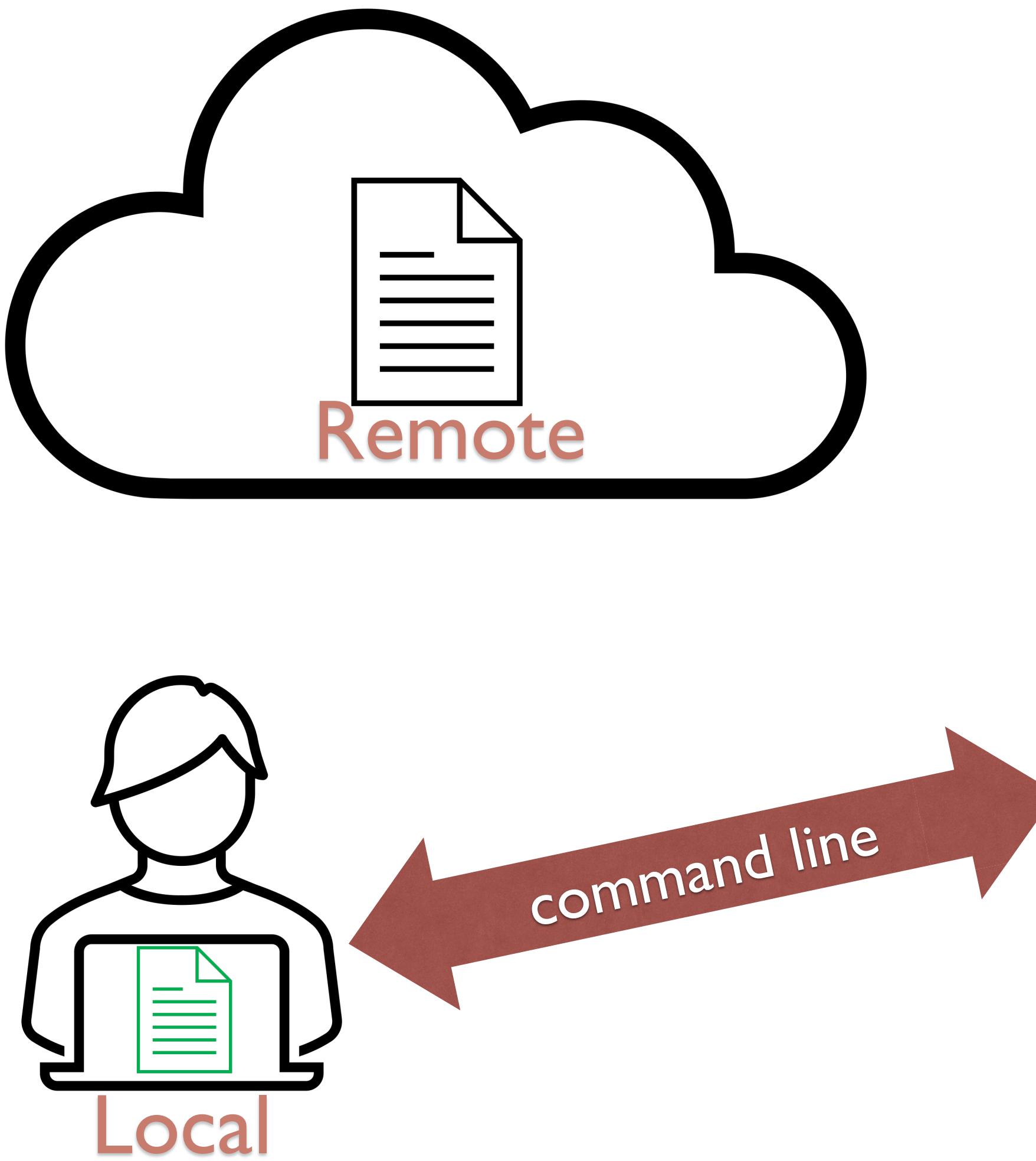
Introduction to Git



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

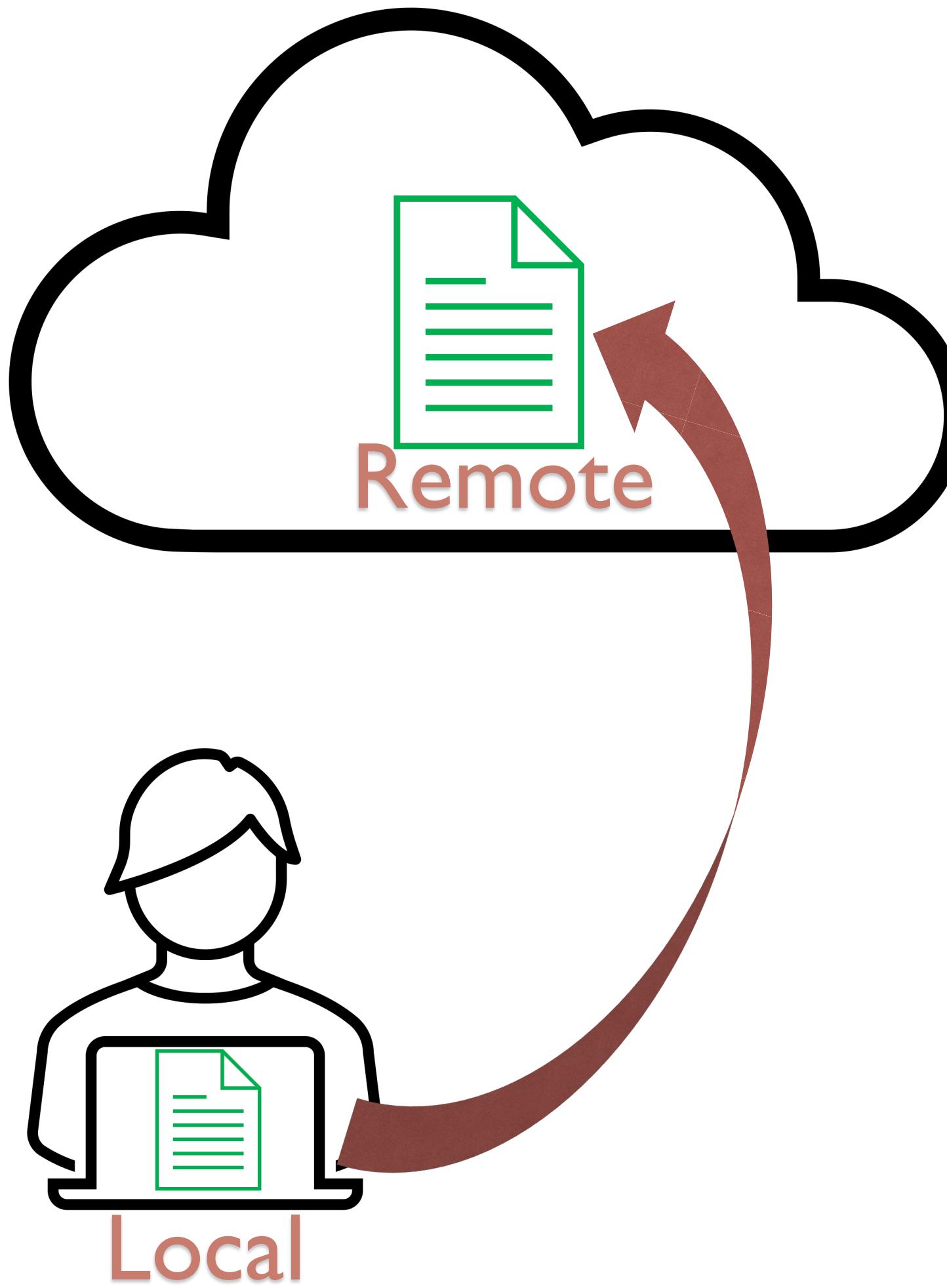


How to use Git?

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 -- 100x47
[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.

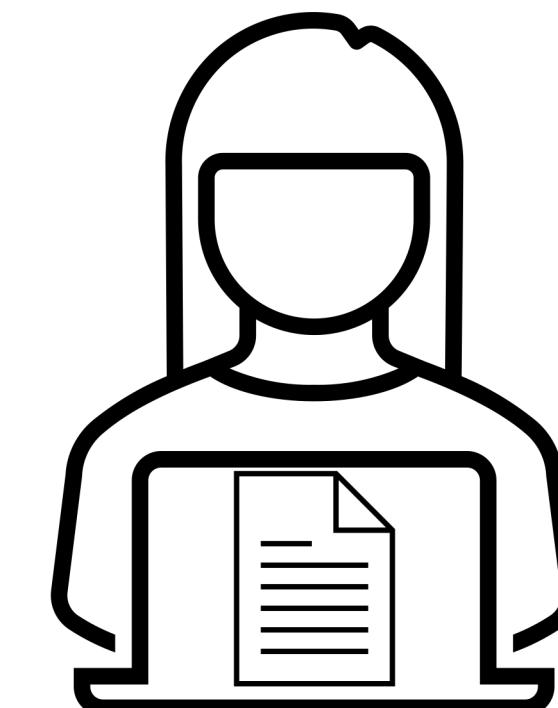
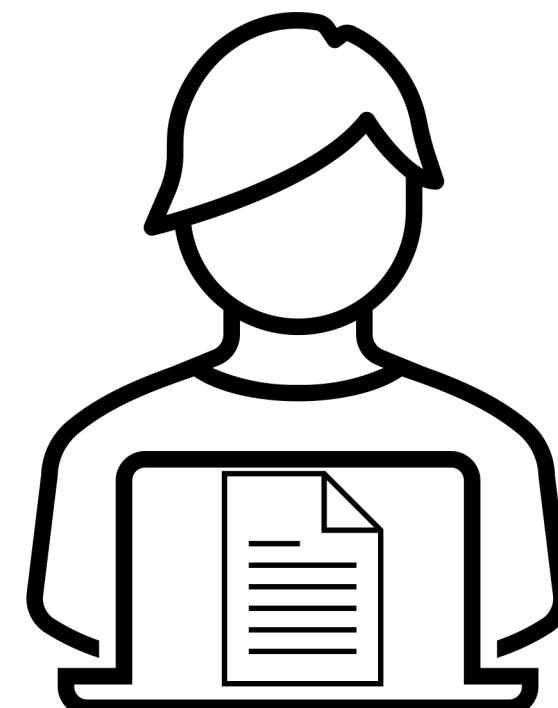
```
ic-01-amosca01 --zsh -- 100x47
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git add .
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git commit -m "part 1 done"
[main ac03563] part 1 done
 1 file changed, 1 insertion(+), 1 deletion(-)
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git push
Enumerating objects: 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

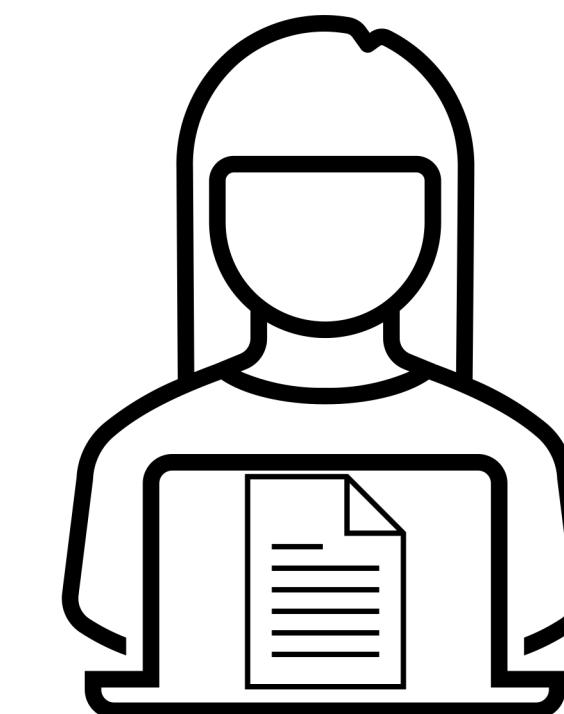
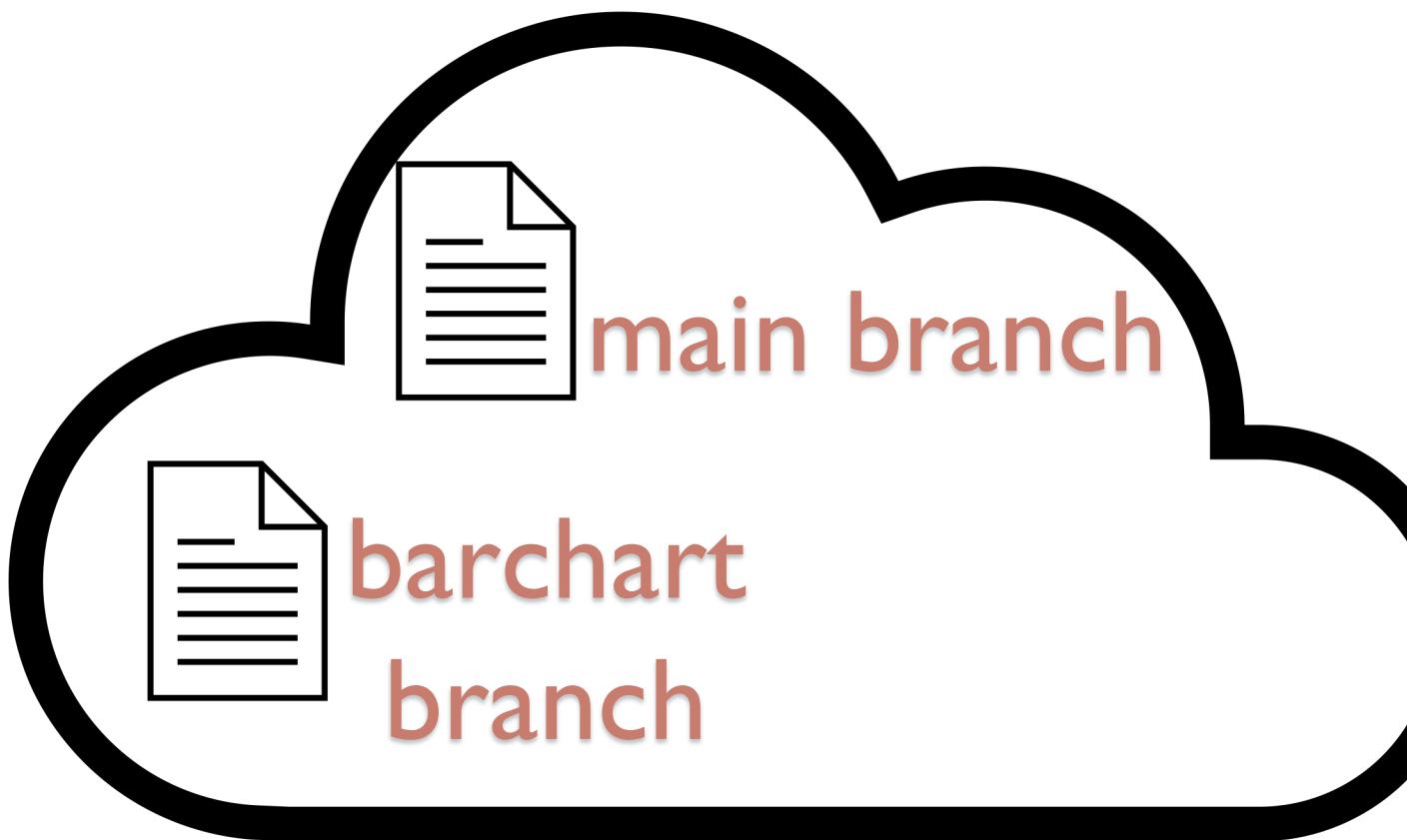


How to use Git?

To collaborate with others, use branches.



Introduction to Git



How to use Git?

To collaborate with others, use branches.

A screenshot of a terminal window titled "ic-01-amosca01 — zsh — 100x47". The terminal shows the following command history:

```
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git branch barchart
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git checkout barchart
Switched to branch 'barchart'
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git branch
* barchart
  main
abmosca@Abs-MacBook-Pro ic-01-amosca01 %
```

Introduction to Git

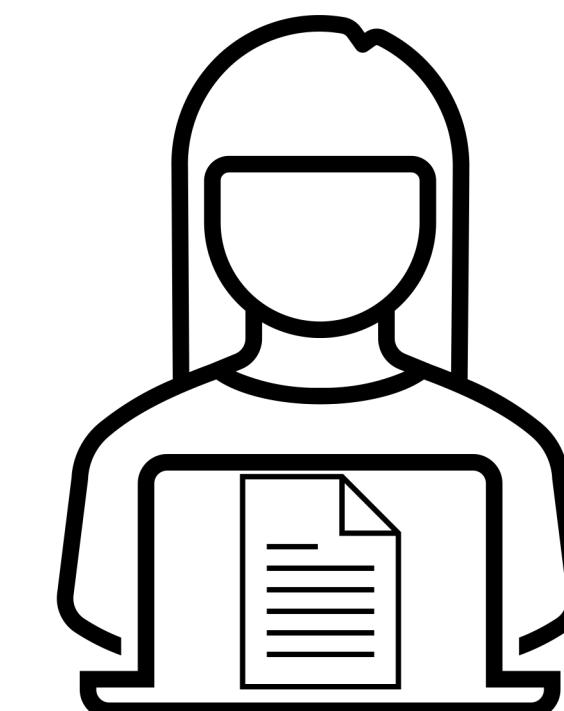
How to use Git?

To collaborate with others, use branches.

```
ic-01-amosca01 - zsh - 100x47
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git branch barchart
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git checkout barchart
Switched to branch 'barchart'
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git branch
* barchart
  main
abmosca@Abs-MacBook-Pro ic-01-amosca01 %
```

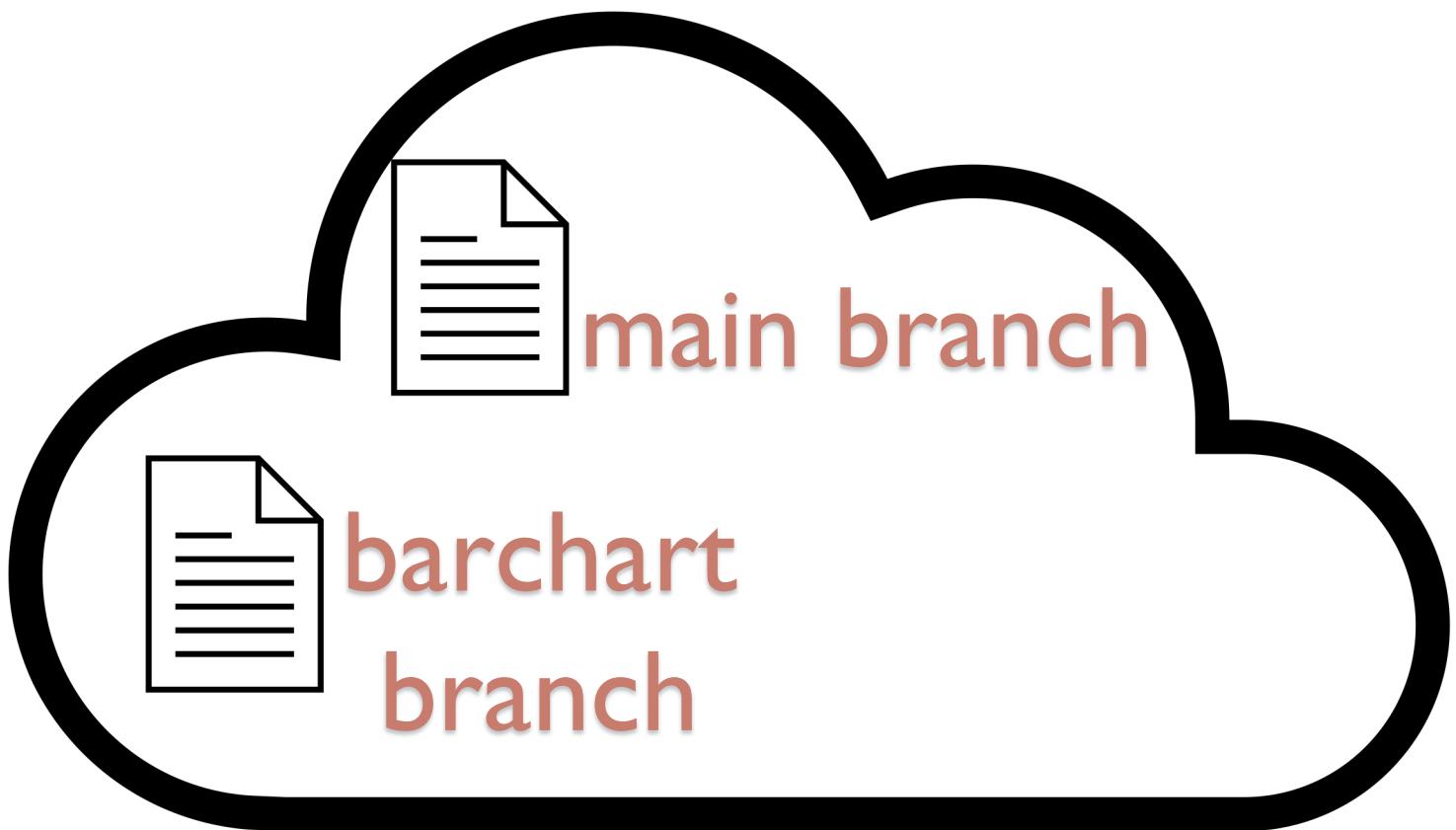


barchart branch



linechart branch

Introduction to Git

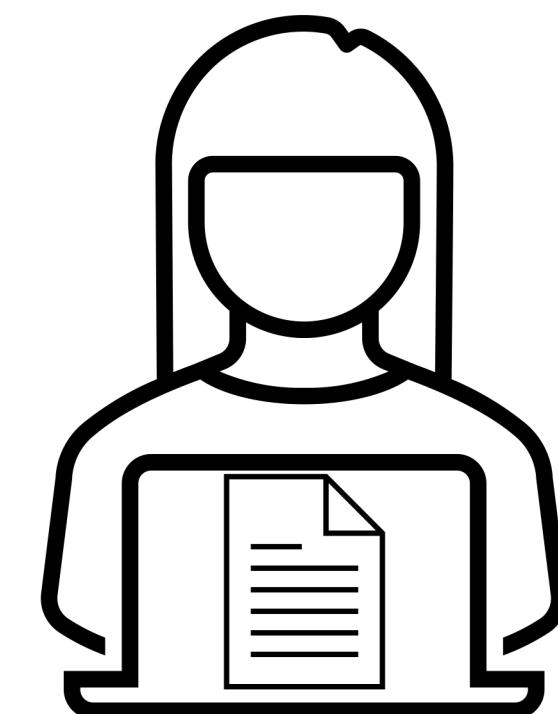


How to use Git?

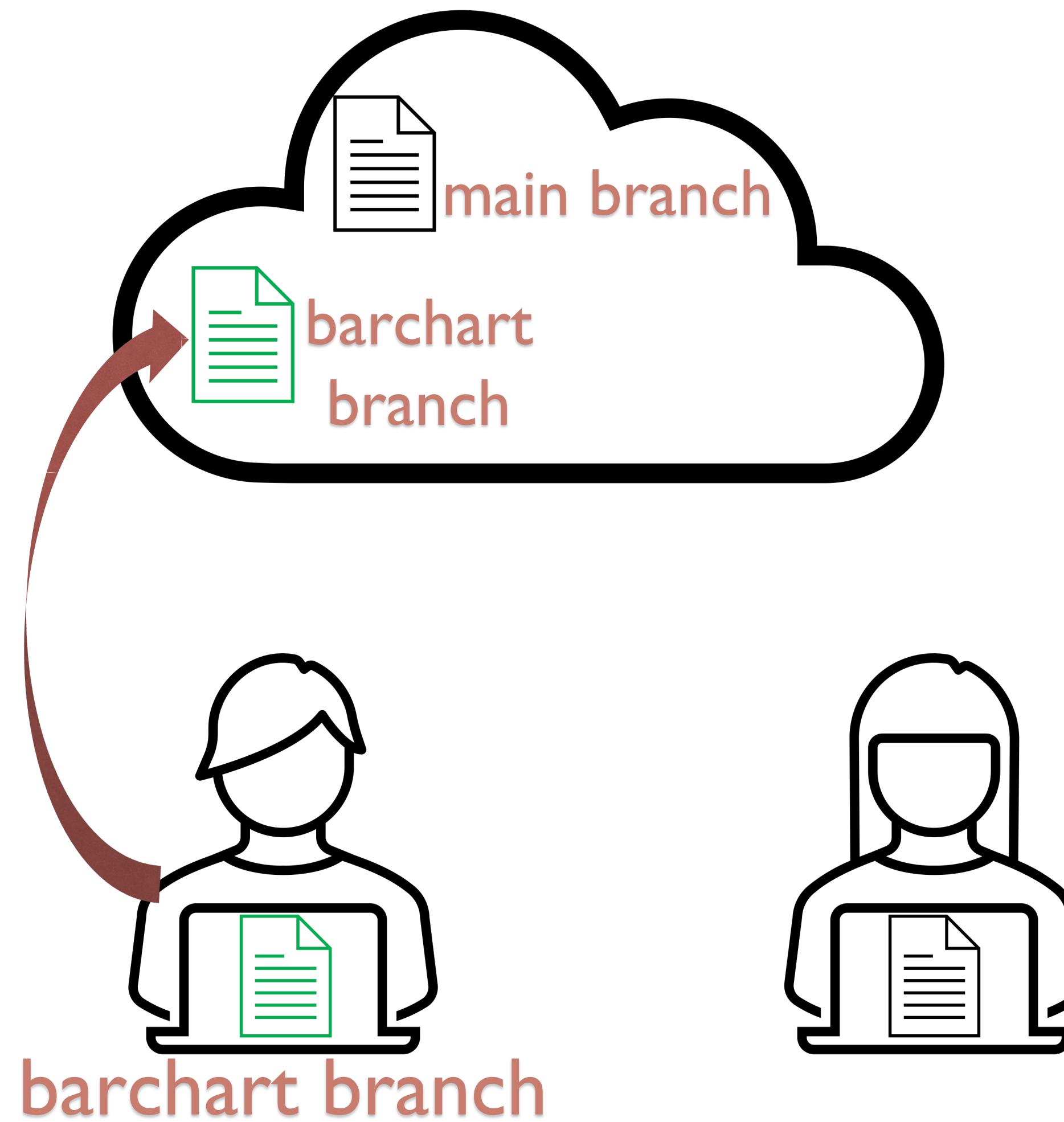
Track changes on your branch.



barchart branch



Introduction to Git

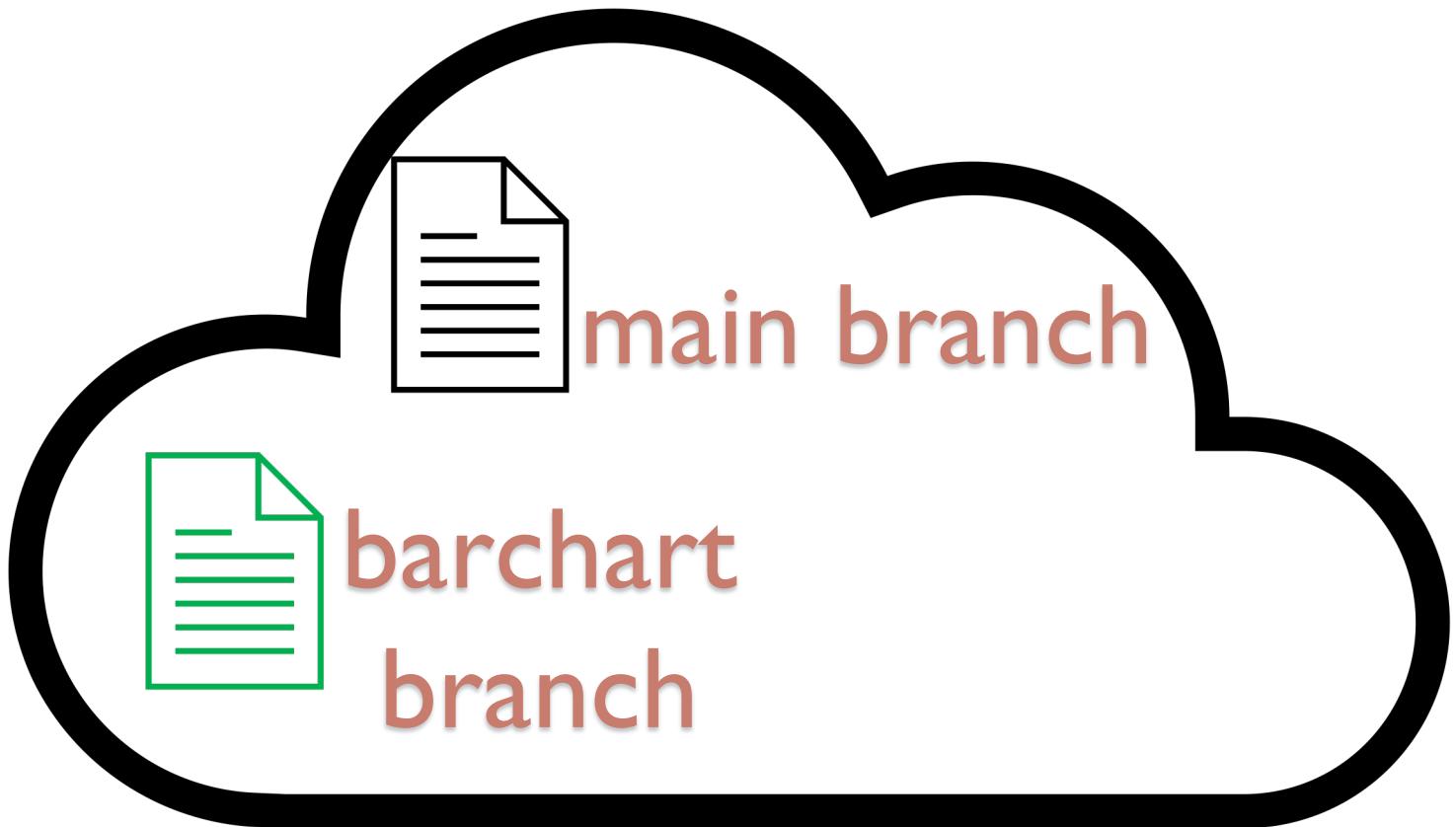


How to use Git?

Track changes on your branch.

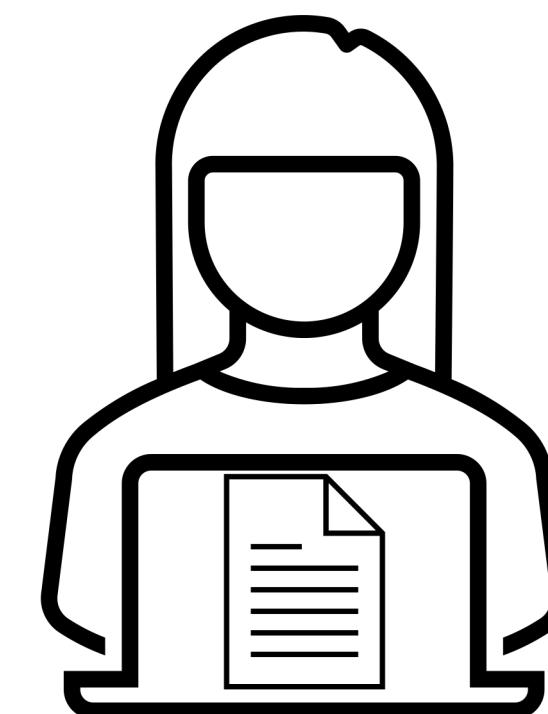
```
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git add .
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git commit -m "edits"
[barchart e45ea28] edits
 2 files changed, 1 insertion(+), 1 deletion(-)
  create mode 100644 .DS_Store
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git push --set-upstream origin barchart
Enumerating objects: 6, done.
Counting objects: 100% (6/6), done.
Delta compression using up to 8 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (4/4), 635 bytes | 635.00 KiB/s, done.
Total 4 (delta 0), reused 0 (delta 0), pack-reused 0
remote:
remote: Create a pull request for 'barchart' on GitHub by visiting:
remote:   https://github.com/DS4200-S22/ic-01-amosca01/pull/new/barchart
remote:
To github.com:DS4200-S22/ic-01-amosca01.git
 * [new branch]      barchart -> barchart
Branch 'barchart' set up to track remote branch 'barchart' from 'origin'.
abmosca@Abs-MacBook-Pro ic-01-amosca01 % ]
```

Introduction to Git

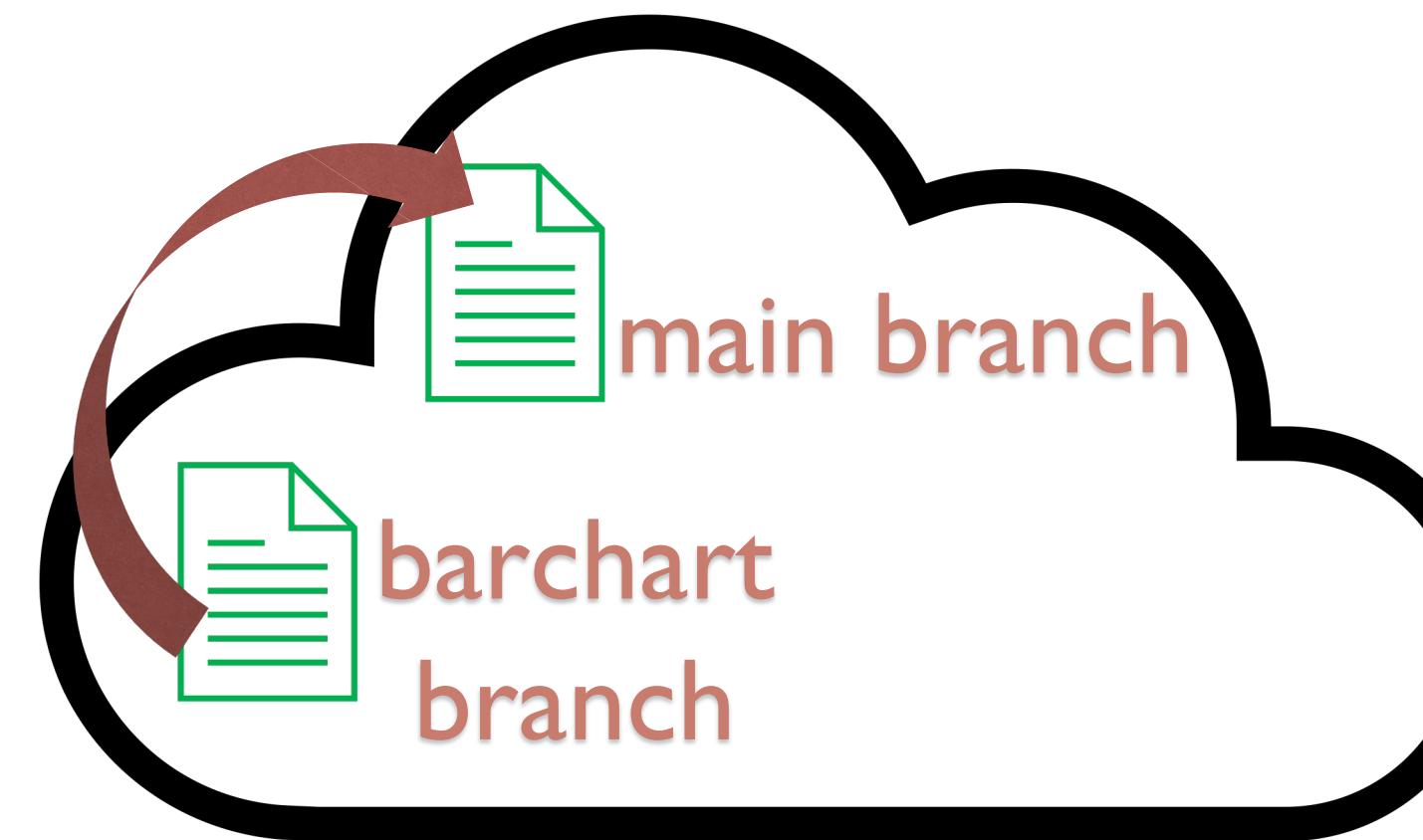


How to use Git?

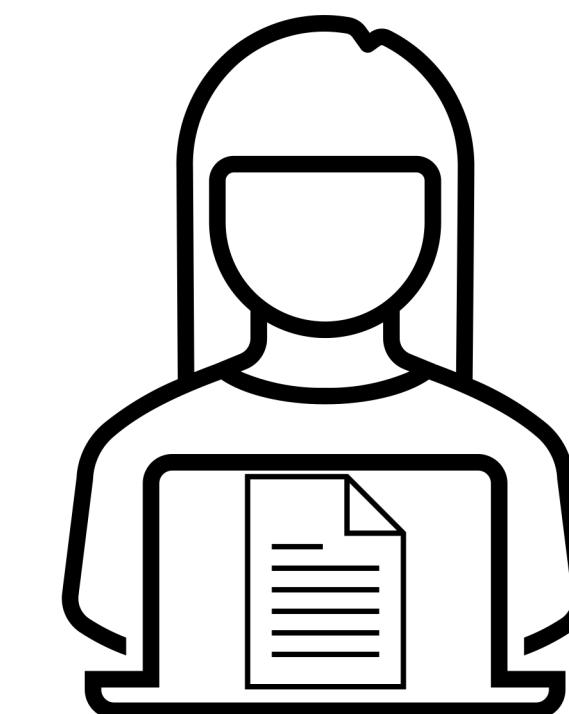
Merge your changes with main to share with teammates.



Introduction to Git



barchart branch



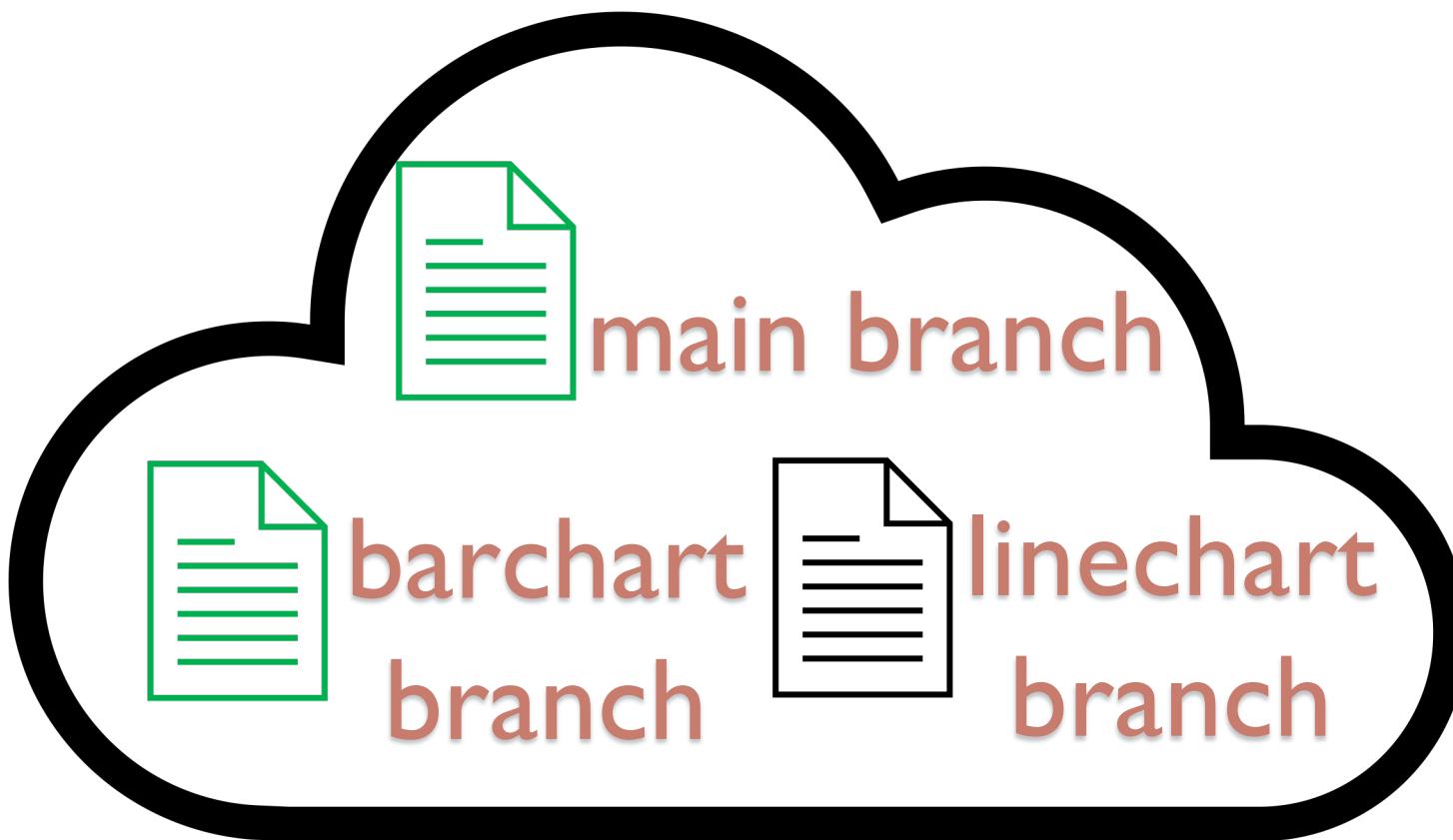
How to use Git?

Merge your changes with main to share with teammates.

A screenshot of a terminal window titled "ic-01-amosca01 — zsh — 100x47". The terminal shows the following command sequence:

```
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git checkout main
Switched to branch 'main'
Your branch is up to date with 'origin/main'.
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git merge barchart
Updating c8bdb0c..4887602
Fast-forward
 .gitignore | 1 + 
index.html | 2 +- 
2 files changed, 2 insertions(+), 1 deletion(-)
create mode 100644 .gitignore
abmosca@Abs-MacBook-Pro ic-01-amosca01 %
```

Introduction to Git

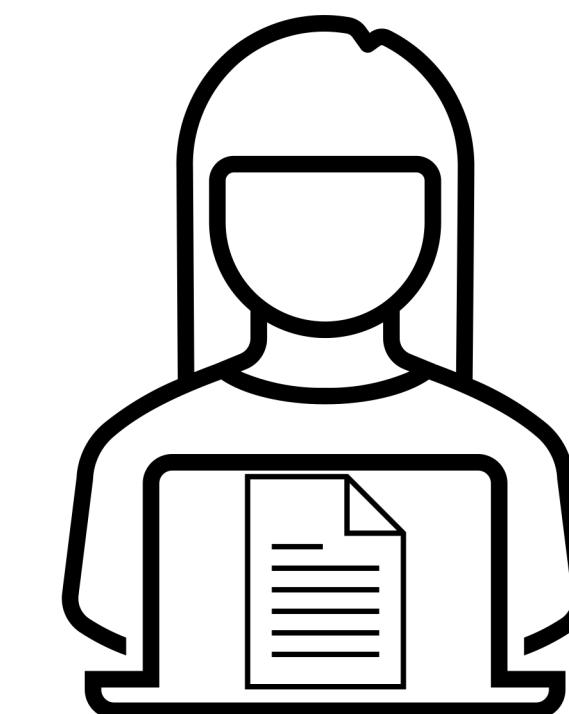


How to use Git?

Pull from main to update your local repo.

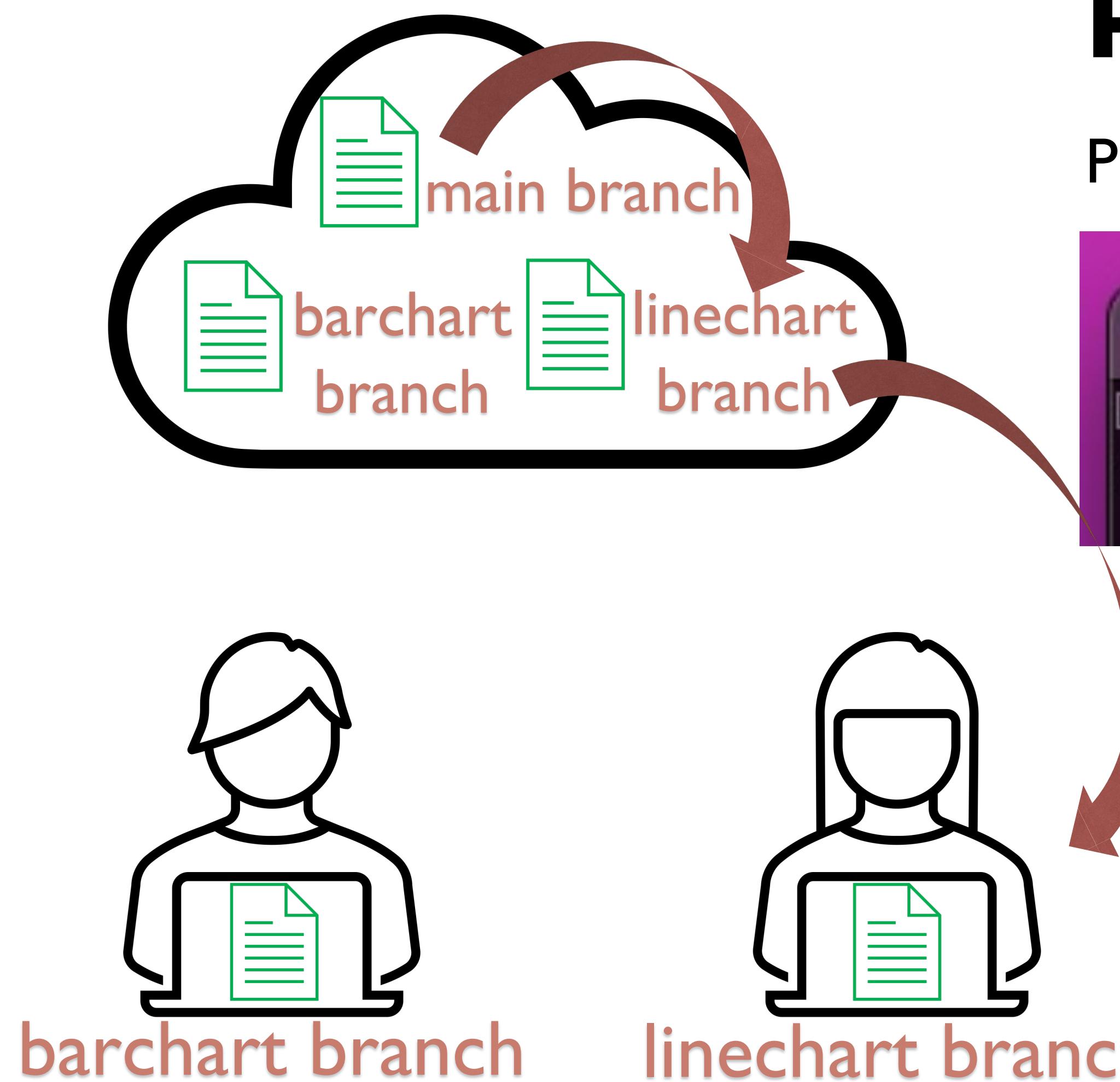


barchart branch



linechart branch

Introduction to Git



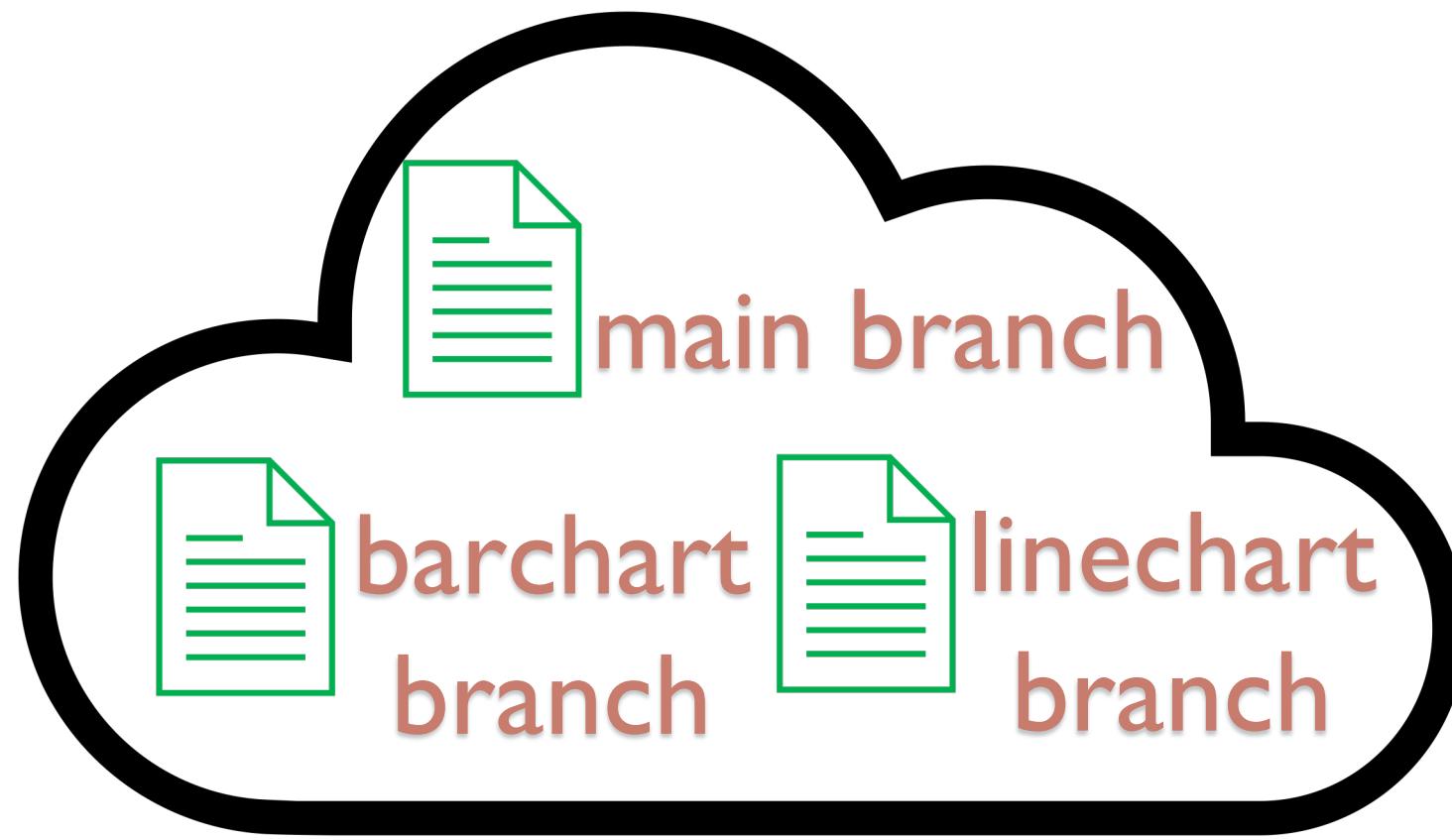
How to use Git?

Pull from main to update your local repo.

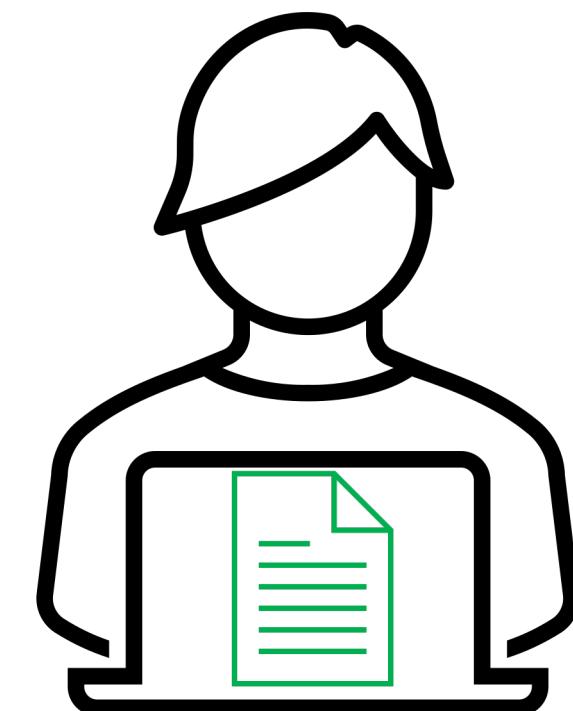
A screenshot of a terminal window titled 'ic-01-amosca01 --zsh-- 100%'. The command entered is 'git pull origin main'. The output shows the command was run from 'github.com:DS4200-S22/ic-01-amosca01'. It indicates the 'main' branch is checked out, and 'FETCH_HEAD' is a symbolic reference to it.

```
[abmosca@Abs-MacBook-Pro ic-01-amosca01 % git pull origin main
From github.com:DS4200-S22/ic-01-amosca01
 * branch      main        -> FETCH_HEAD
```

Introduction to Git



When in doubt, Google!



barchart branch

linechart branch

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**

GitHub in This Class

Let's look at an example with ic-01

→ <https://classroom.github.com/a/pYEPMoF8>

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.