Data storytelling

Reminders

Lab 2 - Due Friday Feb 9 at 2:30pm

Lab 3 - Due Friday Feb 9 at 2:30pm

- On discord there are suggestions about how to label your x-axis and how to make the initial plot using the raw data (line-initial)
- https://colab.research.google.com/github/allegheny-college-cmpsc-105-spring-2024/course-materials/blob/main/notes/20240202_lab03_data_transformation.ipynb

Accept the gradebook repo if not already accepted

- https://classroom.github.com/a/bC_S5GJo

More on Lab 03 - dot to dot ipynb

- What is a matrix?
- What does datamat look like and why??
- How are values accessed from the datamat?
- How are values saved in the datamat?
- How are transformations applied?

```
[ ] datamat
```

```
[ ] datamat[:,(variables == 'rawdata').argmax()].astype(float)
```

```
[ ] datamat[:,rawdata].astype(float)
```

```
[ ] datamat[:,transformation1] = np.log2(datamat[:,rawdata].astype(float))
    datamat[:,transformation2] = 10*datamat[:,transformation1].astype(float)
    datamat[:,transformation3] = np.multiply(datamat[:,sign].astype(float), datamat[:,transformation2].astype(float))
    datamat[:,distance] = 60-datamat[:,transformation3].astype(float)
    datamat[:,transformation4] = datamat[:,transformation3].astype(float)+2*datamat[:,distance].astype(float)
```

Recall the LOs

Learning Outcomes

- Develop hypotheses based on motivating problems and/or observations and identify appropriate data to address hypotheses.
- Identify and describe key elements in different types of data visualizations.
- Use web-based platforms to accurately present data sets through multiple visualizations.
- Use an open-source programming language to compute summary statistics and visualize key patterns in the data.
- Contribute to and present structured, web-based documentation that describes data exploration steps and visualization-based conclusions.

data back story

Sheets, Dashboards

python (ipynb)

markdown blog posts

Grading policy vote

https://forms.gle/XD7zCwEPeMzbhLGt9

60% gatorgrade, 4 checks

70% gatorgrade, 3 checks

other

Data storytelling

Goal

Understand the back story of a visualization

- backtrace the steps from the end to the start
- identify alternative branches from the starting point

Example

Aesthetic Responses to Repetition in Unfamiliar Music
Elizabeth Margulis
Empirical Studies of the Arts 2013 31:1, 45-57
https://journals.sagepub.com/doi/epdf/10.2190/EM.31.1.c

- About repetitions in music
- specifically, how different types of repetitions contribute to liking the music

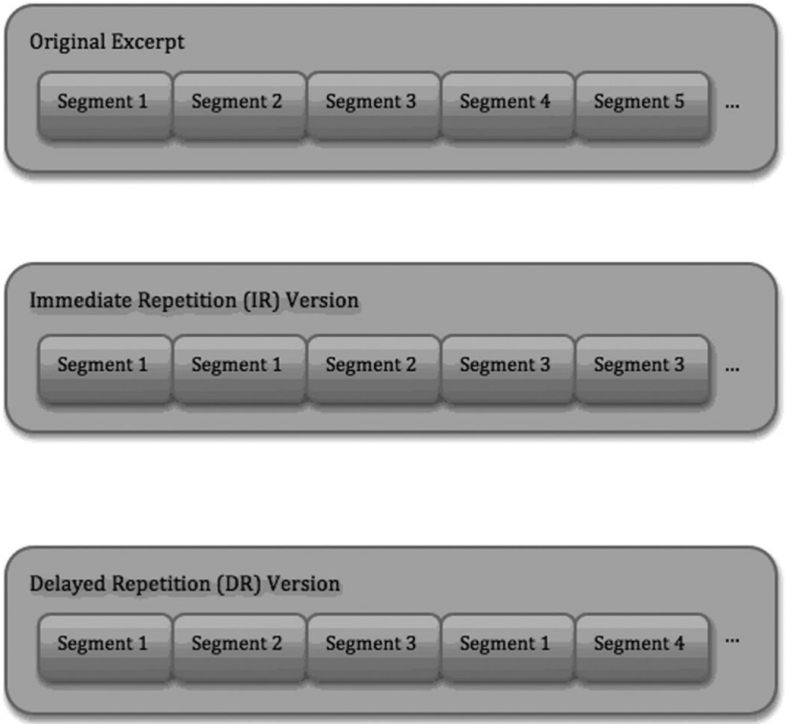
Conditions and Task

Each condition has a different type of repetition

- Original = no repetition
- Immediate repetition
- Delayed repetition

Task

- Rate enjoyment on scale 1-7
- Rate how interesting on scale 1-7
- Rate artistry on scale 1-7



Final plots - Enjoyment

Maximal Enjoyment

Minimal Enjoyment

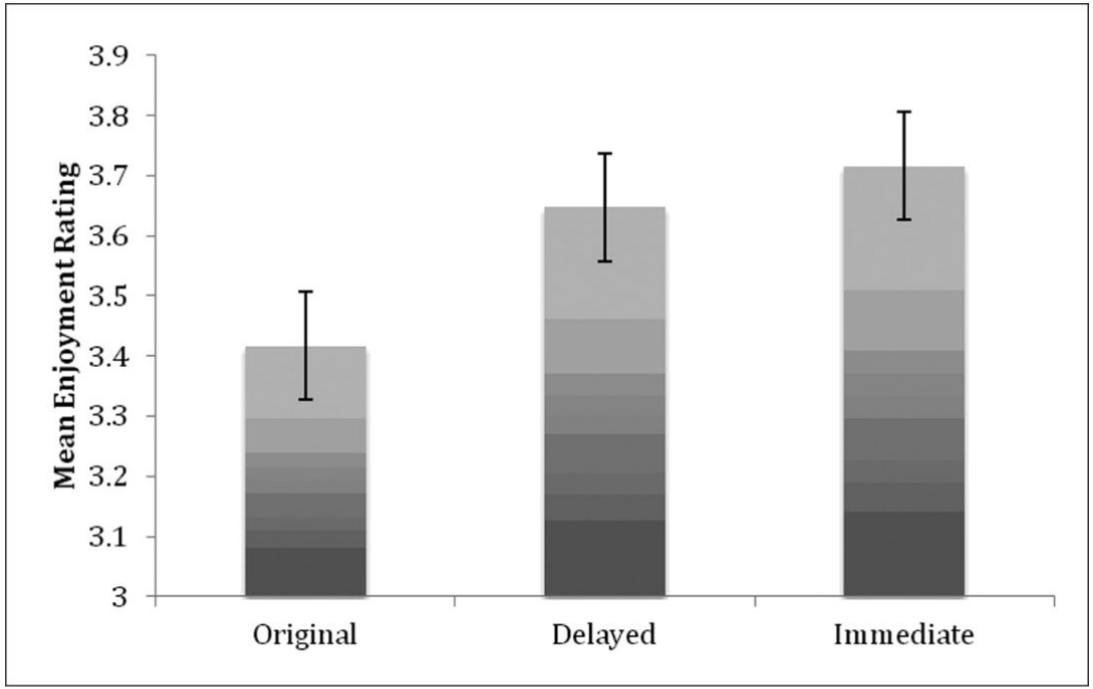
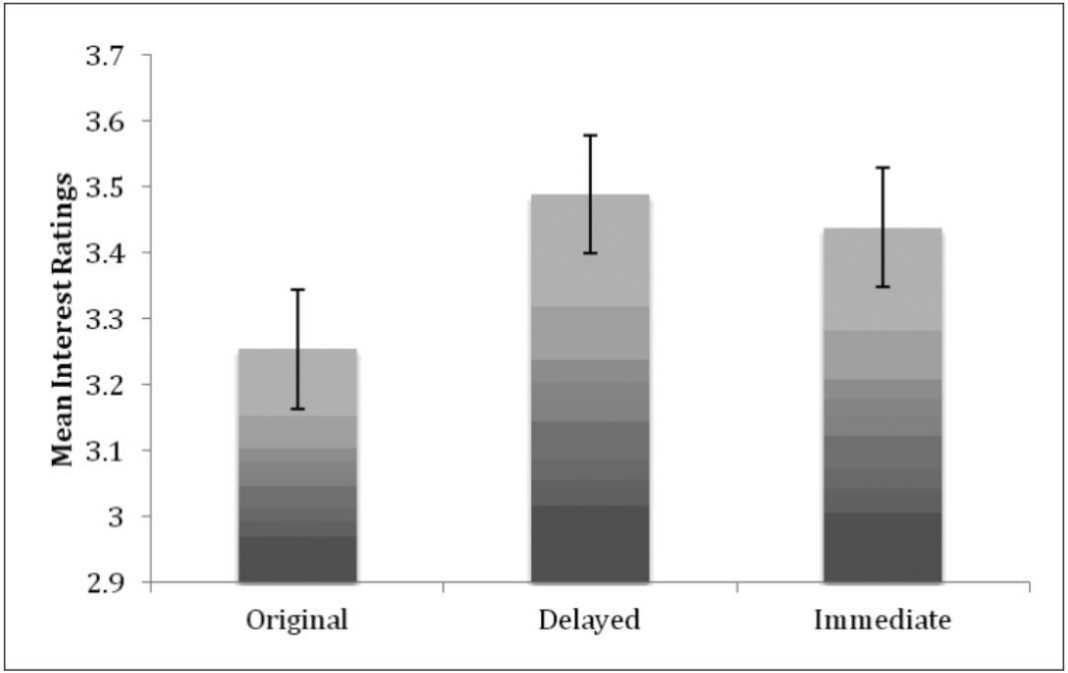


Figure 3. Mean enjoyment rating by Repetition Condition

Final plots - Interest

Maximal Interest



Minimal Interest

Figure 4. Mean interest rating by Repetition Condition

Final plots - Artistry

Written by an artist

Random

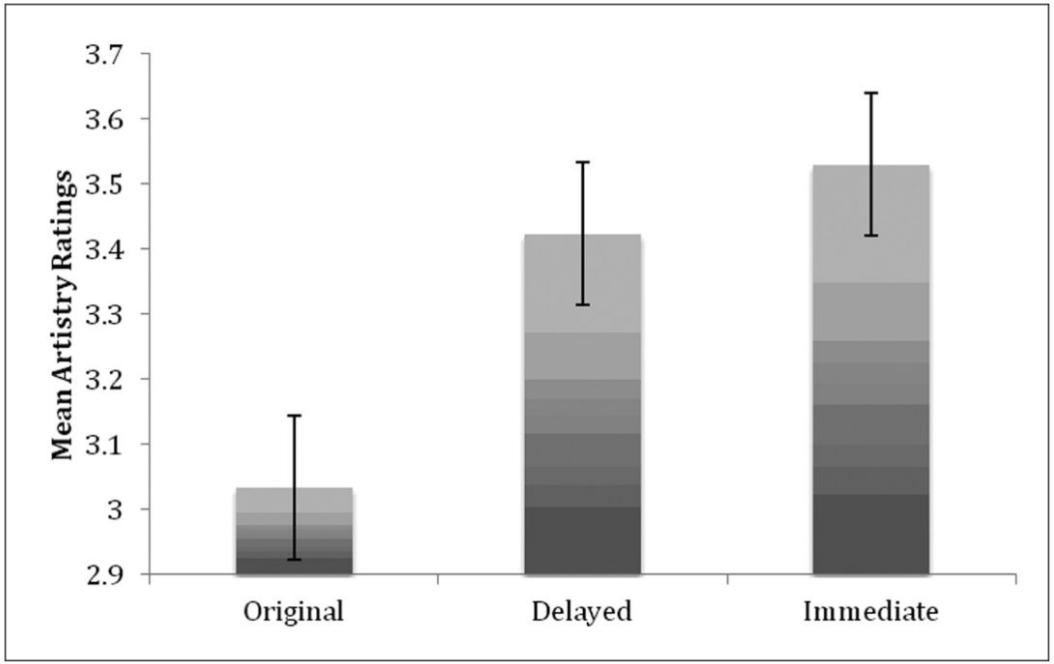


Figure 5. Mean artistry rating by Repetition Condition

How did Margulis get there?

What did the data originally look like?

Start from the end and work backwards

Enjoyment ratings

Interest ratings

Artistry ratings

for every musical excerpt…… (8)

in every repetition condition…… (3)

for every person…….... (33)

~~with different repetition lengths? (only range reported!)~~

Dependent variables

Independent variables

Original Data

subject	excerpt	condition	enjoyment	interest	artistry

Original Data

Order was randomized

subject	excerpt	condition	enjoyment	interest	artistry	trial order

Question

How many variables are there? How many observations are there?

subject	excerpt	condition	enjoyment	interest	artistry	trial order

33 subjects
8 excerpts
3 conditions
3 dependent measures
randomized order

Question

What other final plots could have been made?

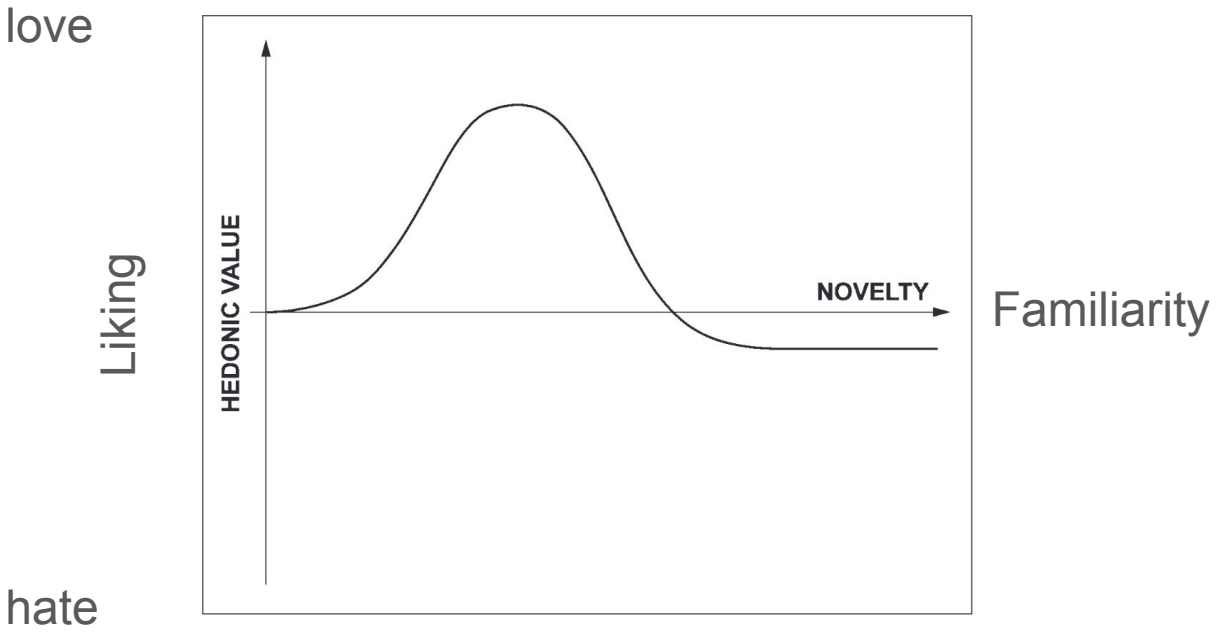
subject	excerpt	condition	enjoyment	interest	artistry	trial order

What insight/knowledge could have been gained from the same data?

Would there have to be any data processing? (transformations or organization?)

The selected pathway through the data was motivated

hypothesis based on prior information:



Group Activity - 4 people per group

Hall of Fame

Go back to the good graphics that you found and suss out the back story.

- What other data had to be gathered?
- What variables had to exist in order to produce the final plots that you see?
- Given those other variables existed, how do you think they arrived at the final plot?
- Where else could exploration have led?
- Could the final plot have been something else?

Installation Continued

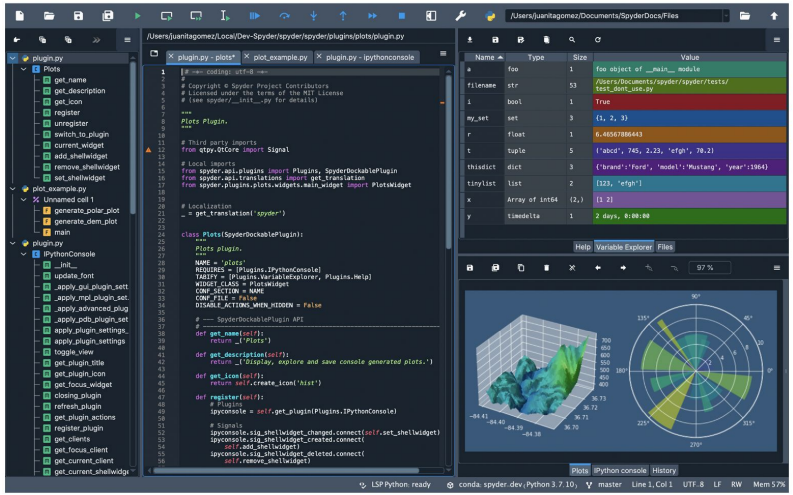
Summary

- Install **spyder** if you don't already have it
- get **SSH key** if you don't already have one
- check for **git** and install if you don't have it

Spyder

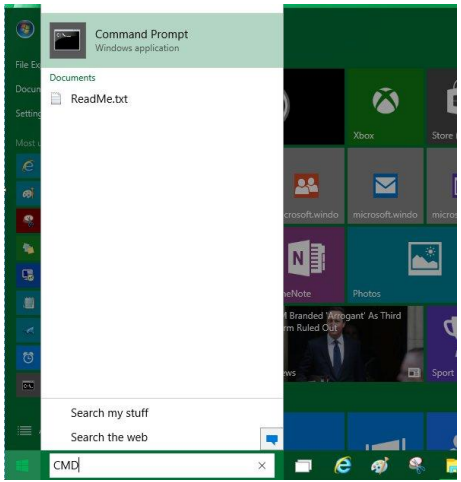
https://www.spyder-ide.org/

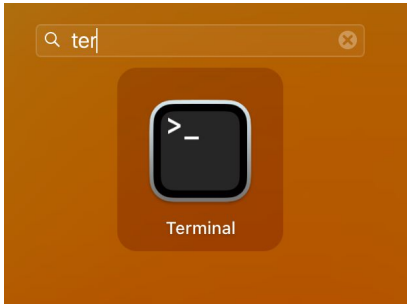
For laptops with windows, linux, macos

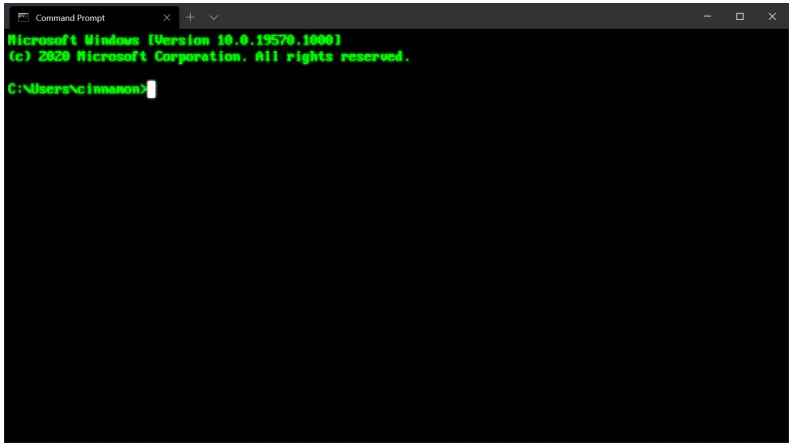


Terminal reminder

Program on your computer to access all files, folders, and programs without using GUI.







What is Git?

https://git-scm.com/downloads

git tracks versions of files so you don't need to do _v01 _v02 in filenames, but old versions can still be found using git

git also is able to transfer your local file changes to GitHub, using the terminal!

To check if you have git, in terminal type: `git --version`

File transfers requires security

Secure connects are often made using passwords

SSH keys are a types of security on top of passwords

- half of the key is public
- half is private

The public key gets stored in GitHub and only allows file transfers if it finds the corresponding private key.

I cannot transfer my local files from my computer to your github using my SSH key!

SSH Key

20240131_colab_intro.ipynb

- In terminal type `ssh-keygen -t rsa -b 4096 -C "youremail@allegheny.edu"`
- hit enter to save in default location
  a. default location is HOME/.ssh/
- enter a passphrase (Don't forget this!!)
- print out the public key by either:
  a. typing `cat ~/.ssh/id_rsa.pub`
  b. opening the public key in notepad or any basic text editor
- copy everything including email
- paste into GitHub: Profile pic > settings > SSH Keys > New Key
- Video on the topic

Hidden files

- Some filenames and directories start with .
- . indicates that the information is hidden
- Like hidden slide in a slide show…it is there, just not showing unless you specifically look.