COGS 108 Week 4 A04/A07

Oct 23, 2023

Slides Adapted from Previous Quarter

slido

How are you doing today?

ⓘ Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

slido

Have you talked with your group mates?

ⓘ Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

AGENDA FOR TODAY

01 LOGISTICS

02 DISCUSSION LAB 3

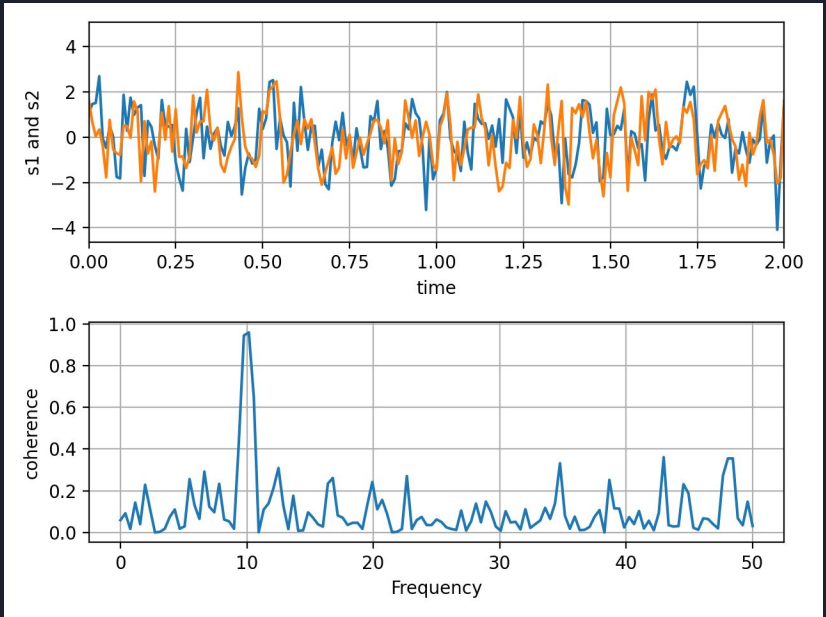
LOGISTICS

DUE DATES

- Quiz 3 is due Oct 23, 11:59PM (Today)
- Project Reviews are due Wednesday Oct 25 11:59PM
- D3 is due Friday, Oct 27, 11:59PM
- Project Proposal is due Nov 1, 11:59PM (next Wednesday)
  - Where do I find this?
    - Go to your Github Repo (Accept your invites!)
    - Find info in ProjectProposalGroup_XXX-Fa23.ipynb

Matplotlib (plt)

- A plotting library for Python
- Makes static, animated, and interactive visualizations in Python.
- Usually imported under the plt alias



https://matplotlib.org/stable/tutorials/index

There are so many ways to make a same plot!

- All of these do the same thing:

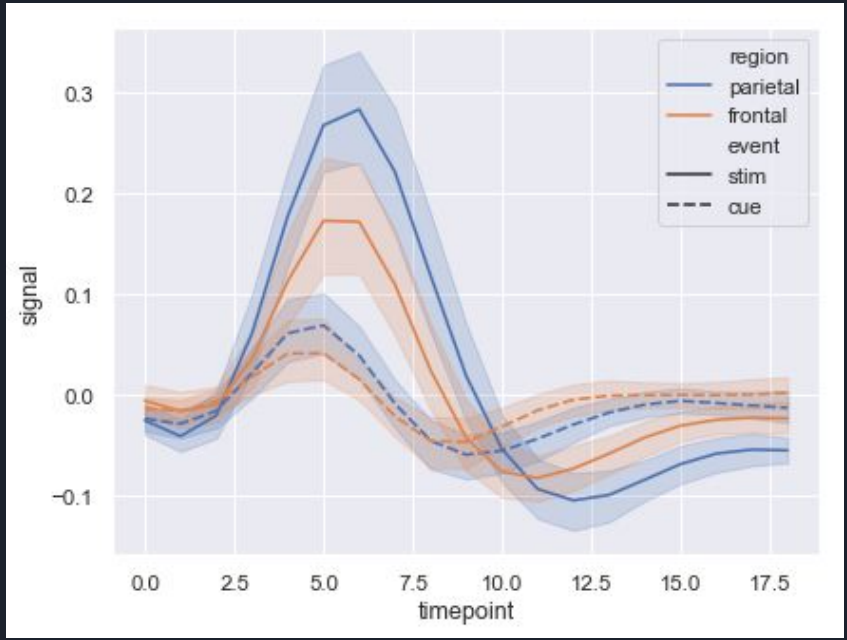
  Line 1: plt.hist(df['income'], 25)
  Line 2: df['income'].hist(bins=25)
  Line 3: df.hist('income', bins=25)

- In Python, most image-based plots created using Matplotlib (plt)
  plt.hist plt.bar plt.plot etc.

- Pandas gives shortcuts for matplotlib plots. Lines 2 and 3 are shortcuts for line 1.

Seaborn

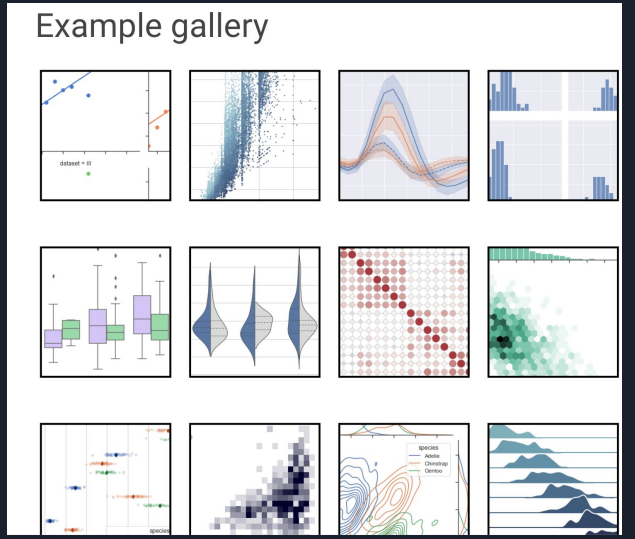
- Makes common statistical charts easy to create, like bar plots with confidence intervals.
- Again, seaborn is really just a bunch of shortcuts for matplotlib.
- Usually imported under the sns alias



https://seaborn.pydata.org/tutorial.html

Tips for Data Visualization

- Google is your best friend.
- E.g. “stacked histogram python”, “scatter plot sns”, “remove legend seaborn”
- Documentation is your friend: https://seaborn.pydata.org/api.html
- Half the time is getting the data in the right format- clean your data



https://seaborn.pydata.org/examples/index.html

Pandas Series and Dataframes

Python3

```python
#importing pandas library
import pandas as pd

#Creating a list
author = ['Jitender', 'Purnima', 'Arpit', 'Jyoti']
#Creating a Series by passing list variable to Series() function
auth_series = pd.Series(author)
#Printing Series
print(auth_series)
```

Output:

```
0    Jitender
1     Purnima
2       Arpit
3       Jyoti
dtype: object
```

One-dimensional ndarray with axis labels (including time series).

Pandas Series and Dataframes

- We have created two lists ‘author’ and article’ which have been passed to Series() functions to create two Series.
- After creating Series, we have created a dictionary and passed Series objects as values of the dictionary and keys of the dictionary will be served as Columns of the dataframe.

Python3

```
#Importing Pandas library
import pandas as pd

#Creating two lists
author = ['Jitender', 'Purnima', 'Arpit', 'Jyoti']
article = [210, 211, 114, 178]
#Creating two Series by passing lists
auth_series = pd.Series(author)
article_series = pd.Series(article)
#Creating a dictionary by passing Series objects as values
frame = { 'Author': auth_series, 'Article': article_series }
#Creating DataFrame by passing Dictionary
result = pd.DataFrame(frame)
#Printing elements of Dataframe
print(result)
```

Output:

```
     Author  Article
0  Jitender      210
1   Purnima      211
2     Arpit      114
3     Jyoti      178
```

DISCUSSION LAB 3 DATA Visualization and Exploratory Data Analysis

THANKS!

Questions on Campuswire or office hours

Office hours: Tue/Thu, 4-5 PM