CSCI 1010 Class 3

Profs. Michael Linderman and Phil Chodrow
Department of Computer Science
Middlebury College

Quiz 1

- 15 minutes (10:00AM-10:15AM). Wait until 10:00AM to begin.
- The quiz is closed everything (no notes, no Internet)
- Place your devices including your phone, laptop, smartwatch, etc. in your bag/backpack, and place your bag at the front or side of the room.
- Place your completed quiz in the box at the front of the room when you are done.
- You are welcome to step out of the room, grabbing your bag on the way, if you finish early, but please do so quietly.

`plt.figure` creates a new figure for plotting. When in the plotting workflow would `plt.figure(figsize=(10, 6))` need to invoked?

a. Before creating any plots on that figure

b. After creating plots but before showing them

c. Any time before `plt.show()`

d. It can be called anytime

Slide 2 Notes

Answer: A

Creating or resizing a figure must be done before any plots are added to it.

You have data showing daily website traffic for multiple marketing campaigns over 90 days. You are interested in visualizing trends over time across campaigns and device types (mobile vs desktop). Which plot type would be most appropriate for this task?

a. Scatter plot

b. Line plot

c. Histogram

d. Stacked bar chart

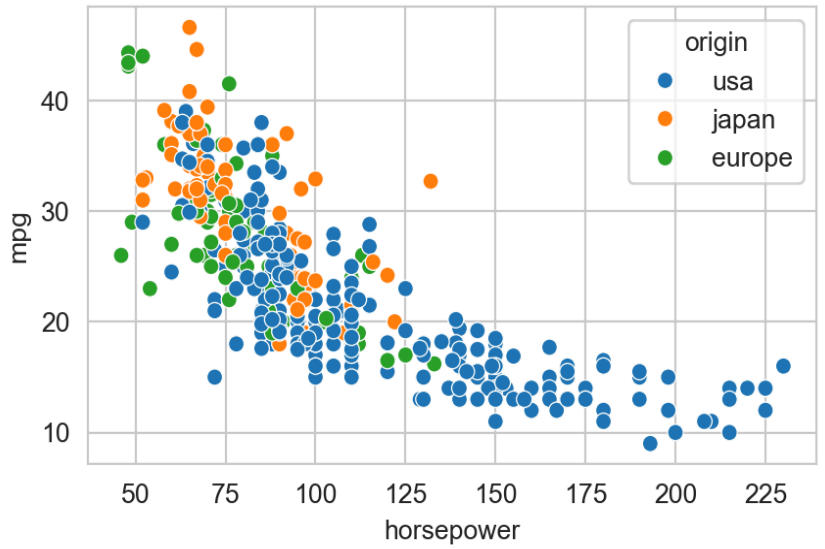
Slide 3 Notes

Answer: B

Since it is data over time (time-series) data, points are connected. Thus a line plot is most appropriate. In contrast it would be more difficult to observe trends over time with a scatter plot. A stacked (or grouped bar chart) would have similar challenges for observing trends over time. A histogram is for visualizing distributions.

```
1 mpg.dtypes
```

```
mpg             float64
cylinders         int64
displacement    float64
horsepower      float64
weight            int64
acceleration    float64
model_year        int64
origin           object
name             object
dtype: object
```



Which code snippet would produce this plot?

D

a. `sns.scatterplot(mpg, x="horsepower", y="mpg", style="origin")`

b. `sns.scatterplot(mpg, x="mpg", y="horsepower", hue="origin")`

c. `sns.scatterplot(mpg, x="horsepower", y="mpg", style="origin", hue="origin")`

d. `sns.scatterplot(mpg, x="horsepower", y="mpg", hue="origin")`

Slide 4 Notes

Answer: D

Pay close attention to the variable mapping for x, y and hue (and the constant `style`).

	country	year	type	value
0	Afghanistan	1999	cases	745
1	Afghanistan	1999	population	19987071
2	Afghanistan	2000	cases	2666
3	Afghanistan	2000	population	20595360
4	Brazil	1999	cases	37737

Which of the following best describes whether this data frame could be used to plot per-capita diseases over time for each country?

a. No, it is missing necessary data.

b. No, because values in some columns are duplicated.

c. Yes, but it would need to grouped/reshaped first.

d. Yes, it could be used as is.

Slide 5 Notes

Answer: C

All the relevant data is present, but awkwardly structured for plotting. We would need to reshape it to create columns for cases and population or, somewhat equivalently, group on country and year. I would not describe this as “tidy data”, as the value column has two different variables: cases and population.