

STAT107 Data Science Discovery

LAB: PLOTS

Man Fung (Heman) Leung

Spring, 2022

University of Illinois at Urbana-Champaign

- Please work in a group of 2–4 students
 - collaboration is important in data science!
 - meet new friends and discuss :)
 - let us know if you have any questions
- **Attendance form**
 - you can come up if you do not want to use this form
 - submit before you leave the lab
- Midterm 1
 - average GPA question removed
 - wish you all the best!

Random fact of the day

The **Zen of Python** can be displayed by entering `import this`.

- Glad to see some of you followed my tips
- Check email for score decomposition
- 1.4/2.3/3.1/3.5: missing reflection(s)
- 2.1/3.3: if the dataframe is not randomized, then `head()` or `tail()` is wrong
- 2.1: sample from `df` twice (with `random_state`) for treatment and control is wrong. Shuffle treatment and control from part 1 is also wrong
- 3.1: reasonable answers are accepted but not shoe size. You cannot block on the variable of interest
- 3.2: if you do blocking in 3.3, I let you go this time
- 3.3: stratified random sampling is biased (thus wrong) if not 6 from `block1` and 5 from `block2`

- [Main page](#)
- Hints:
 - pre-lab/post-lab survey each for 2 extra points
 - 3.1 approach 1: use `import matplotlib.pyplot as plt` and call `plt.show()` after each box plot; see [this post](#)
 - 3.1 approach 2: use `.groupby("Gender").plot.box()`
- Submit your work. Feel free to:
 - ask us questions
 - leave whenever you finish the lab