Module 2: Data Visualization with R

Jeffrey Woo

MSDS Program, University of Virginia

Welcome

- You can download this set of slides. Find with the materials for the live session in Module 2.
- Remind me to record the live session!
- Recommended: put yourself on mute unless you want to speak.
- There is a "raise hand" button for you. Click on "Reactions" in the panel at the bottom.

Agenda

- A few logistical comments.
- Q&A
- Some comments on Module 2.
- Small group discussion of guided question set
- Large group discussion of guided question set plus other questions that pop up

Getting help...

- Office hours on Zoom: Mondays and Thursdays (GTA).
- Discussion forums.
- Email is to be used for more personal questions.
- We can also set up time to meet on Zoom if a discussion is needed.
- For coding help, show me your entire code / provide me your R script. Adding comments to provide some info about what you're trying to do helps. I need to reproduce your error on my end. Just giving me the line of code preceding the error doesn't help me help you. I want to be able to help you in a way that is consistent with your code.

HW Logistics...

- HWs are due every Monday following the live session.
- My advice: aim to complete HW by Friday so you have time to prep for the upcoming module. So you can ask me questions during Monday's office hours and to give yourself time to go over the material rather than rush through the material.
- The GTA should be grading HWs within a week. You should receive an email notification when he is done.
- Solutions posted on Tuesday morning. You may want to run the script for HW 1 solution to get the correct data frame needed for HW 2.

- Any questions?
- Logistical or from module 2?

ggplot2 Package

- The functions we are using to create visualizations are from the ggplot2 package.
- Loading the tidyverse package automatically loads a number of other packages, including ggplot2 and dplyr.
- The visualizations from ggplot2 are a lot nicer than similar functions in base R.

Loading Packages

Be aware that as you load more and more packages, conflicts in function names can arise.

- Notice there is a function called filter() in the dplyr package as well as the stats package, which is part of base R.

- Typing in filter() will use the function that comes from the package that was most recently loaded. So the function from the dplyr package is used.

- To use the function from a specific package, type stats::filter().

Data Visualization in R

We have covered the basics. A whole class can be dedicated to this topic.

- Be sure to give descriptive titles and labels for axes, especially as you generate more visualizations (for example in projects).

Some Conventions...

When we have two quantitative variables and we create a
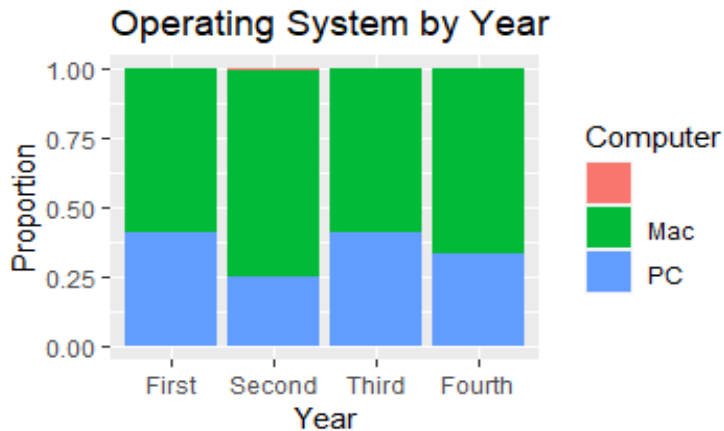scatterplot:

- **Response** variable on y-axis, **predictor** variable on x-axis.
- Can be called a scatterplot of response variable **against**
  predictor variable.
- When we create a regression model, we regress response
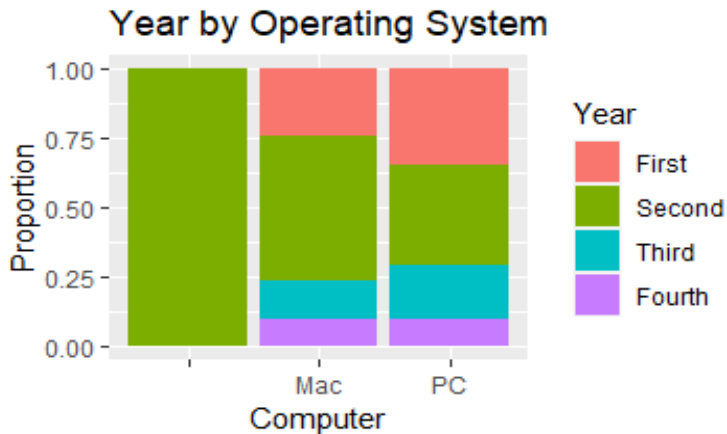 variable **against** predictor variable.

Some Conventions...

When we have two categorical variables and we create a barchart:

- With ggplot, put response variable as "fill".

Operating System by Year

Small Group Discussion

- Materials can be found under Module 2 Live session.
- Have the guided question set and corresponding data set open.
- Have R open.
- Recommended: have easy access to your notes, textbook, as well as the tutorial.
- You can see who your group members are. As well as some roles you will have in your small group. Roles will rotate each session.
- You will be evaluating your group members based on how prepared they are and their contribution to the small group discussion.

What is due...

- Homework 2, due Sep 12.
- Submit via assignments
- Please check the guidelines for homeworks.
- Get prepared for next Tuesday's live session.
- You should have completed reviewing Module A.

WARNING: Collab does not work well when you open multiple tabs. If you upload your HW with multiple tabs open, your submission is likely to screw up! You can always check if the submission went through correctly, and you can resubmit if needed.