LAB 02

PH142 Fall 2025

Announcements

- **Lab02:** due 9/5 at 11:59pm

- **Quiz01:** due 9/5 at 11:59pm

- Course Drop deadline (without fee): 9/5 at 11:59pm

Week 2 Lecture Review

Exploring Data

- Get to know your data before analyzing!

- Key functions:
    - head(data): view the first few rows
    - dim(data): view the number of rows and columns
    - names(data): view column names
    - str(data): view the types of variables

Week 2 Lecture Review

dplyr Review (P1)

Function	Purpose	Input	Output	Notes
rename( )	Rename variables	rename(old_dataset, new_name = old_name)	Renamed dataset	Can do multiple variables at one!
select( )	Select a subset of variables	select(dataset, column1, column2)	Dataset with specific columns	Can remove variables with -
arrange( )	Sort by a variable	arrange(dataset, column)	Dataset arranged by specific columns	Can sort by multiple variables, default is ascending order, descending order with -

Week 2 Lecture Review

dplyr Review (P2)

Function	Purpose	Input	Output	Notes
filter()	Select a subset of rows	filter(dataset, condition)	Dataset with rows based on conditions	==, <, >, <=, >=, != Or (|), and(, &)
mutate()	Add new variables to a dataset	mutate(dataset, new_column = data)	Dataset with new variable	You can call existing variables in the dataset to create your new variable
group_by	Group data by a categorical variable	group_by(column)	Dataset grouped by variable	Use with summarize()!
summarize()	Summarize a statistic	summarize(statistic _name = calculation)	Tibble with grouped data and statistic for each group	Use with group_by()!

Week 2 Lecture Review

Pipe Operators

Key operator: Pipe (%>%)

- Connects functions together so you can do more than one in a single code chunk
- Shorthands your code by only calling the dataset once
- Read as "and then"... R will read your functions from left to right

Template: dataset %>% function1( ) %>% function2( )...

Week 2 Lecture Review

ggplot

Template: ggplot(data= dataset, aes(x=var1, y=var2)) + geom_point( )

- geom_line( )
- geom_histogram( )
- geom_bar( )

Other Changes:

- + labs(title=" " , y= " ", x=" ")
- aes(col= variable1, lty=variable2)

Week 2 Lecture Review

Types of Variables

<u>Categorical variables:</u> variables that have grouping levels

- Nominal variables: have no underlying order or rank
- Ordinal variables: can be ordered or ranked

<u>Quantitative variable:</u> continuous, numeric variables that that you can perform mathematical operations on
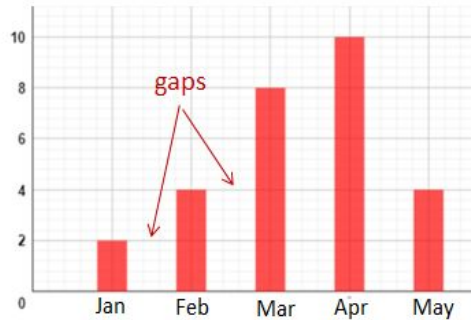
- Discrete variables: can be counted
- Continuous variables: can be measured precisely

Week 2 Lecture Review

Visualizing Distributions

Bar Charts (categorical variables):
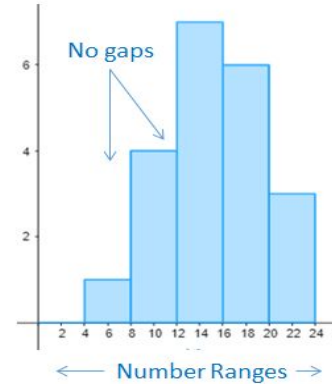- Y-axis: a count/percentage of each category

Histograms (quantitative variables):
- Distribution of values across bins
- Describe shape, center, spread, outliers

Bar Chart

Histogram

LAB 02 Walkthrough

Lab Submission

- Follow the directions on the LAB02 file
- Submit using the **Terminal Tab** (next to the console in the bottom left pane)
  - Copy and paste the given line into the terminal
  - Follow prompts (NOTE: the terminal will **not** show your password being typed out!)
- **<u>CHECK IN GRADESCOPE THAT ALL YOUR TESTS PASSED</u>**