STAT 534: Homework 1

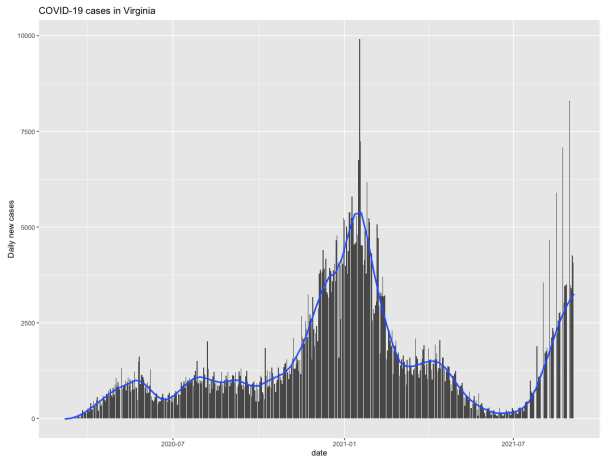
Fall 2021

Due: Monday, September 13, 12:30 pm

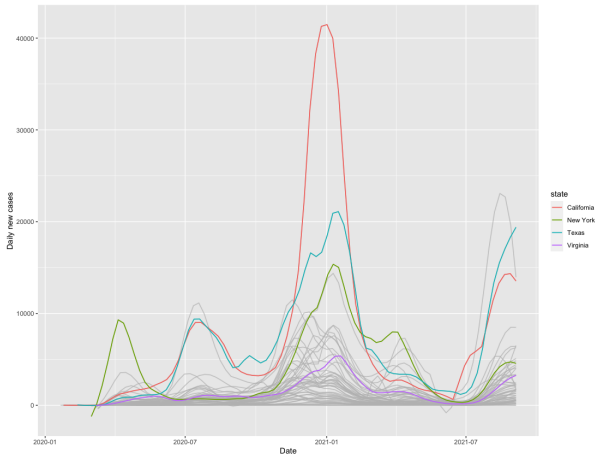
1. (Excerpted from Problem 3.8 in textbook) Using the *penguins* data set from the *palmerpenguins* package: (Hint: use na.omit(penguins) to remove cases with missing values.)
   (a) Create a scatterplot of bill_length_mm against bill_depth_mm where individual species are colored and a regression line is added to each species. What do you observe about the association of bill depth and bill length?
   (b) Repeat the same scatterplot but now separate your plot into facets by sex. How would you summarize the association between bill depth and bill length?

2. Using *storm* data from the *dplyr* package:
   (a) Produce a histogram of the pressure variable. Fill your bars using the category variable.
   (b) Repeat part (a) with the wind speed variable.
   (c) Use `geom_path()` to plot the path of each tropical storm in 2014. Use color to distinguish the storms from one another. Which storm in 2014 made it the furthest North?
   (d) Challenge: show changes in the category variable along the paths in part (c). (Hint: group the data by names and add another aesthetic that maps category.)
   Remarks: The group aesthetic is by default set to the interaction of all discrete variables in the plot. This choice often partitions the data correctly, but when it does not, or when no discrete variable is used in the plot, you will need to explicitly define the grouping structure by mapping group to a variable that has a different value for each group.

3. COVID-19 vignette
   (a) Use data in *us_cases.txt* (data source: New York Times repository of COVID-19 data) to reproduce Figure (a) with `geom_col()` and `geom_smooth(span = 0.1, method = "loess", se = F)`.
   (b) Challenge: reproduce Figure (b) with `gghighlight()` function in the *gghighlight* package.

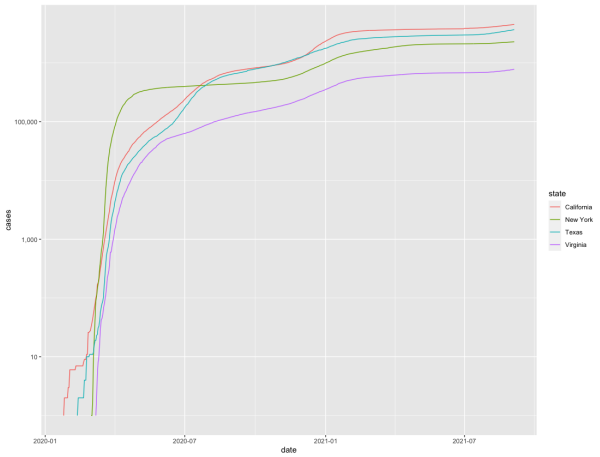


(a)



(b)

(c) When tracking a disease, the rate of growth is particularly important, and is proportional to the logarithm of the case count. Reproduce the figure below with `scale_y_log10(labels = scales::comma)`.



(d) Use data in *vaccine.txt* (data source: Johns Hopkins repository of COVID-19 data) to create an informative graphic that demonstrates the vaccination rollout across the US at the current stage. You may follow examples at Johns Hopkins coronavirus resource center or create your own plot.