Exploratory Data Analysis and Visualisation 11374/11517

Week 3 Lab Exercises

1. The "trees" dataset from the *datasets* package in R provides some information on trees and their measurements
    a. Read the help description of the dataset (?trees)
    b. Create a histogram of the *Height* variable with suitable axis labels, title etc., a theme of your choice and a binwidth of 10.
    c. Are there any better binwidth options? What could these reveal? Should we take caution in any action?

2. Consider the "cabbages" data frame from the MASS package in R, which gives weights and vitamin C content of cabbage heads.
    a. Read the help description about the "cabbages" dataset.
    b. Create histograms of the "HeadWt" and "VitC" variables in R. Describe the histograms in terms of their distributions, their implications etc.
    c. Compare the two histograms – are they similar in pattern? Perhaps not? Do tell.
    d. Wrangle the dataset and by looking at the tibbles created, generate some insight. Though this method is not visual, it encourages further plots (more on this next week). Consider grouping the dataset and using the summarise function to compare certain variables of interest, such as Vitamin C content and head weight. You could use the mutate function as well.
    e. Create bar plots showing the frequency of both the *Cult* variable and the *Date* variable. What do you notice in terms of their frequencies and more specifically, the levels of each variable?
    f. Based on your observations, could you wrangle the data more comprehensively than in part d.?

3. Recall the "airquality.csv" file from the Week 2 Lab. It contains daily air quality measurements in New York, May to September 1973 on the following variables:
    - Ozone: Mean ozone in parts per billion from 1300 to 1500 hours at Roosevelt Island
    - Solar.R: Solar radiation in Langleys in the frequency band 4000–7700 Angstroms from 0800 to 1200 hours at Central Park
    - Wind: Average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport
    - Temp: Maximum daily temperature in degrees Fahrenheit at La Guardia Airport

    a. Brainstorm some potential problems of interest you may want to answer using this dataset. Consider different ways in which you can group the data.
    b. Create charts which help answer your brainstormed problems of interest.