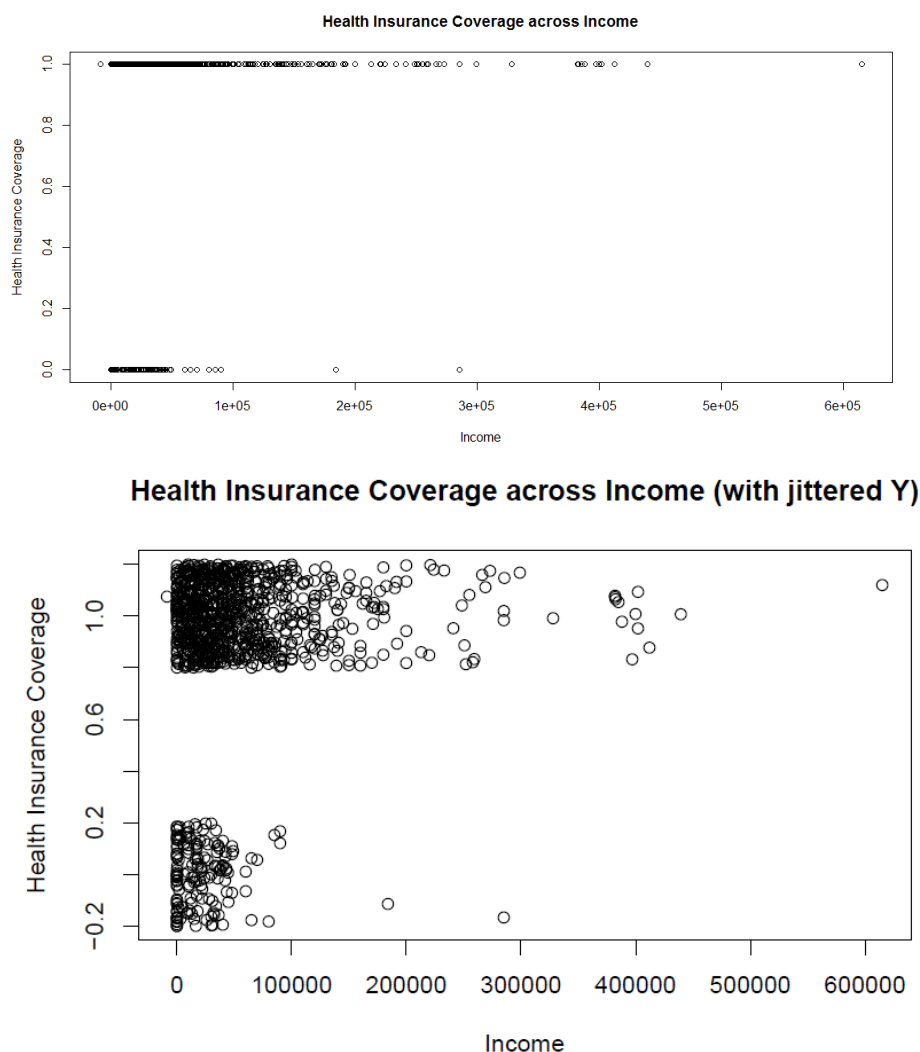


Exercise 3.1 Solution

Part 1. Concepts

Q3.1. In unit 2, R exercise, we learnt that the **cor()** function handles missing values via the optional argument **use =**. But from `?cor` documentation, there are several argument values that you can choose. The following 3 options are popular: (1) `complete.obs`, (2) `pairwise.complete`, (3) `NA.or.complete`. Explain, in your own words, what is the difference between (1) and (2)? Note: There is a subtle but very important difference.

Q3.2. By using "health_ins_cust.csv," draw the following scatter plots of health insurance coverage and income without and with jittered function. Discuss what you can see from the graphs.



Q3.3 The following graph shows the scatterplot of Age and Income. Here, we do not need to use jitter. Why?



Part 2. Run R Script

R script file: ADA1.3.2 flights14.R

Dataset: flights14.csv

Objectives

- Compare data.frame (df) in Base R approach vs data.table approach

Run ADA1.3.2 flights14.R

- Run one line of code at a time via <CTRL> + <Enter> in windows or <CMD> + <Enter> in Apple, or "Run" button.
- So that you can see the results of that specific line and compare against different approaches.
- Do not run the entire script by clicking "Source" button as it will run from first line to last line of code and you will not be able to see effects line by line.