


HW2

1.1

- 1) Year, sex, and name are categorical variable
- 2) Count is quantitative variable
- 3) Taffy is the value of a case

1.2

president	in office	number of states
Lincoln, Abraham	 1861	it depends
Lincoln, Abraham	1862	it depends
Lincoln, Abraham	1863	it depends
Lincoln, Abraham	1864	it depends
Lincoln, Abraham	1865	it depends
George Washington	1791	16
George Washington	1792	16
George Washington	1793	16
George Washington	1794	16
George Washington	1795	16
George Washington	1796	16
George Washington	1797	16
George Washington	1798	16
George Washington	1799	16
Martin Van Buren	1837	26
Martin Van Buren	1838	26
Martin Van Buren	1839	26
Martin Van Buren	1840	26
Martin Van Buren	1841	26

1.3

- 1) A: Year, Algeria, Brazil, Columbia
B: Country, Y2000, Y2001
C: Country, Year, Value

2) A: the years B: the countries C: The country in a year

1.4

- 1) There are five variables
- 2) Cases are the data for each airline, airport, flight, plane, and weather
- 3) Airline, airport, flight, plane, and weather are all categorical variable

1.5

- 1) The sex age and gun are variables
- 2) There is only one jane poole ran in 2004 at age 56, no additional data are needed

2.1

```
aphide=read.csv("aphideData.csv",header = T)
##Assign the data to aphide

plot(aphide[,1:3], pch = 20)
##plot function

data.frame(aphide)
##data table

##2
install.packages("corrplot")
##package

nyRestCorMat=cor(aphide[,1:3])

library(corrplot)
##library function

corrplot(nyRestCorMat, method = "number")
##plot function
```

There contain assignment, function, data table and package.

2.2

Only the file name should be wrapped with quotation mark

2.3

The Current Population Survey (CPS) is used to supplement census information between census years. These data consist of a random sample of persons from the CPS85, with information on wages and other characteristics of the workers, including sex, number of years of education, years of work experience, occupational status, region of residence and union membership.

2.4

The statement should be: `help(package, package<- "NHANES")`

2.5

1) : another is `require()`

2) : the name of the function after `detach()` is `search()`

2.6

3 "MyData" is wrong, no quotation

5 small sample is wrong, no blank

9 `sqrt()` is wrong, it is a function

2.7

- 1) There are 19 variable
- 2) There are 336776 cases
- 3) The cases means the data for each variable such as month, year, dep_time and so on.
- 4) Year, month, day, dep_time, sched_dep_time, dep_delay, arr_time, sched_arr_time, arr_delay, carrier, flight, tailnum, origin, dest, time_hour are categorical variable; air_time, distance, hour, minute are quantitative variable.
- 5) air_time is minutes, distance is mile