Introduction to R

Catherine Holloway and Mine Dogucu

January 29, 2018

Projects

- Make a folder on your computer and name it R Ladies.
- File > New Project > New Directory > New Project Write Introduction to R as Directory Name and Browse to select the R Ladies folder. Cretae Project
- File > New Script

Opening Datasets

Download **North Carolina Births** and **Gun Violence** datasets from

https://www.openintro.org/stat/data/

Save these datasets in your Introduction to R folder.

Opening Datasets

```
nc <- read.csv("Introduction to R/nc.csv")
print(nc)
```

Summarizing datasets : str()

```
str(nc)
```

```
## 'data.frame':    1000 obs. of  13 variables:
##  $ fage          : int  NA NA 19 21 NA NA 18 17 NA 20 ...
##  $ mage          : int  13 14 15 15 15 15 15 15 16 16 ...
##  $ mature        : Factor w/ 2 levels "mature mom","younger mom": 2 2 2 2 2 2 2 2 2 2 ...
##  $ weeks         : int  39 42 37 41 39 38 37 35 38 37 ...
##  $ premie        : Factor w/ 2 levels "full term","premie": 1 1 1 1 1 1 1 2 1 1 ...
##  $ visits        : int  10 15 11 6 9 19 12 5 9 13 ...
##  $ marital       : Factor w/ 2 levels "married","not married": 1 1 1 1 1 1 1 1 1 1 ...
##  $ gained        : int  38 20 38 34 27 22 76 15 NA 52 ...
##  $ weight        : num  7.63 7.88 6.63 8 6.38 5.38 8.44 4.69 8.81 6.94 ...
##  $ lowbirthweight: Factor w/ 2 levels "low","not low": 2 2 2 2 2 1 2 1 2 2 ...
##  $ gender        : Factor w/ 2 levels "female","male": 2 2 1 2 1 2 2 2 2 1 ...
##  $ habit         : Factor w/ 2 levels "nonsmoker","smoker": 1 1 1 1 1 1 1 1 1 1 ...
##  $ whitemom      : Factor w/ 2 levels "not white","white": 1 1 2 2 1 1 1 1 2 2 ...
```

Summarizing datasets : head()

```
head(nc)
```

```
##   fage mage      mature weeks    premie visits marital gained weight
## 1   NA   13 younger mom    39 full term     10 married     38   7.63
## 2   NA   14 younger mom    42 full term     15 married     20   7.88
## 3   19   15 younger mom    37 full term     11 married     38   6.63
## 4   21   15 younger mom    41 full term      6 married     34   8.00
## 5   NA   15 younger mom    39 full term      9 married     27   6.38
## 6   NA   15 younger mom    38 full term     19 married     22   5.38
##   lowbirthweight gender     habit  whitemom
## 1        not low   male nonsmoker not white
## 2        not low   male nonsmoker not white
## 3        not low female nonsmoker     white
## 4        not low   male nonsmoker     white
## 5        not low female nonsmoker not white
## 6            low   male nonsmoker not white
```

Summarizing datasets : tail()

```
tail(nc)
```

```
##      fage mage     mature weeks    premie visits     marital gained weight
## 995    NA   41 mature mom    33    premie     13 not married      0   5.69
## 996    47   42 mature mom    40 full term     10 not married     26   8.44
## 997    34   42 mature mom    38 full term     18 not married     20   6.19
## 998    39   45 mature mom    40 full term     15 not married     32   6.94
## 999    55   46 mature mom    31    premie      8 not married     25   4.56
## 1000   45   50 mature mom    39 full term     14 not married     23   7.13
##      lowbirthweight gender     habit  whitemom
## 995         not low female nonsmoker not white
## 996         not low   male nonsmoker not white
## 997         not low female nonsmoker     white
## 998         not low female nonsmoker     white
## 999             low female nonsmoker not white
## 1000        not low female nonsmoker     white
```

Descriptive Statisitics

```
summary(nc)
```

```
##       fage            mage            mature        weeks
##  Min.   :14.00   Min.   :13   mature mom :133   Min.   :20.00
##  1st Qu.:25.00   1st Qu.:22   younger mom:867   1st Qu.:37.00
##  Median :30.00   Median :27                     Median :39.00
##  Mean   :30.26   Mean   :27                     Mean   :38.33
##  3rd Qu.:35.00   3rd Qu.:32                     3rd Qu.:40.00
##  Max.   :55.00   Max.   :50                     Max.   :45.00
##  NA's   :171                                    NA's   :2
##        premie        visits            marital        gained
##  full term:846   Min.   : 0.0   married    :386   Min.   : 0.00
##  premie   :152   1st Qu.:10.0   not married:613   1st Qu.:20.00
##  NA's     :  2   Median :12.0   NA's       :  1   Median :30.00
##                 Mean   :12.1                     Mean   :30.33
##                 3rd Qu.:15.0                     3rd Qu.:38.00
```

```
##                  Max.   :30.0                     Max.   :85.00
##                  NA's   :9                       NA's   :27
##      weight       lowbirthweight    gender          habit
##  Min.   : 1.000   low    :111    female:503   nonsmoker:873
##  1st Qu.: 6.380   not low:889    male  :497   smoker   :126
##  Median : 7.310                                NA's     :  1
##  Mean   : 7.101
##  3rd Qu.: 8.060
##  Max.   :11.750
##
##       whitemom
##  not white:284
##  white    :714
##  NA's     :  2
##
##
##
##
```

Selecting Variables

```
mean(nc$mage)
```

```
## [1] 27
```

Arguments

Can you find mean age of fathers?

Arguments

```
mean(nc$fage)
```

```
## [1] NA
```

Arguments

```
mean(nc$fage, na.rm=FALSE)
```

```
## [1] NA
```

Using R help

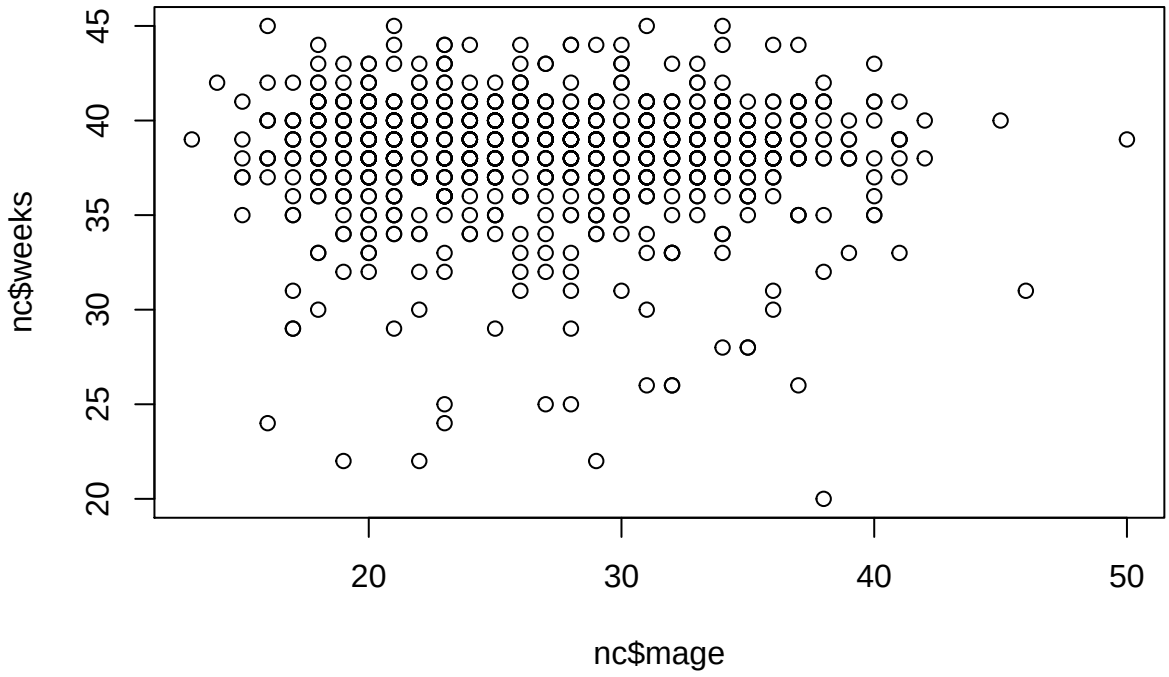
Let's assume we are interested interested in making a scatterplot to show the relationship between _mother's age and *mage* and *weeks* and we do not know how to do this yet.

Using R help

```
?plot
```

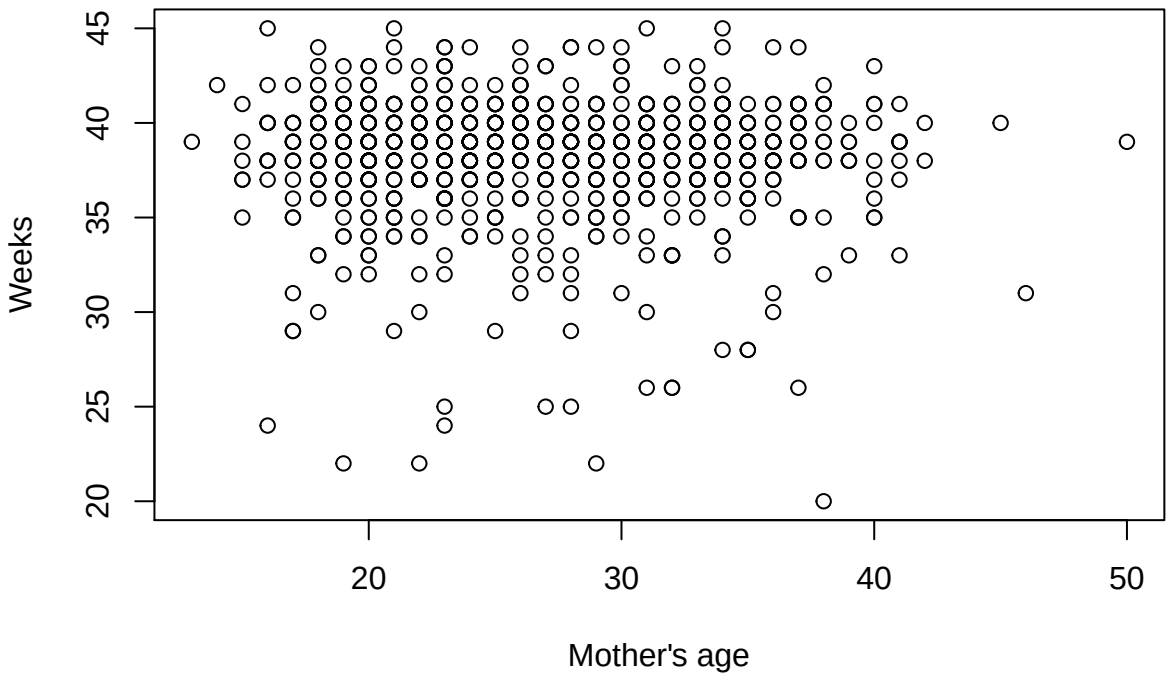
Using R help

```
plot(nc$mage, nc$weeks)
```



Using R help

```
plot(nc$mage, nc$weeks, xlab="Mother's age", ylab="Weeks")
```

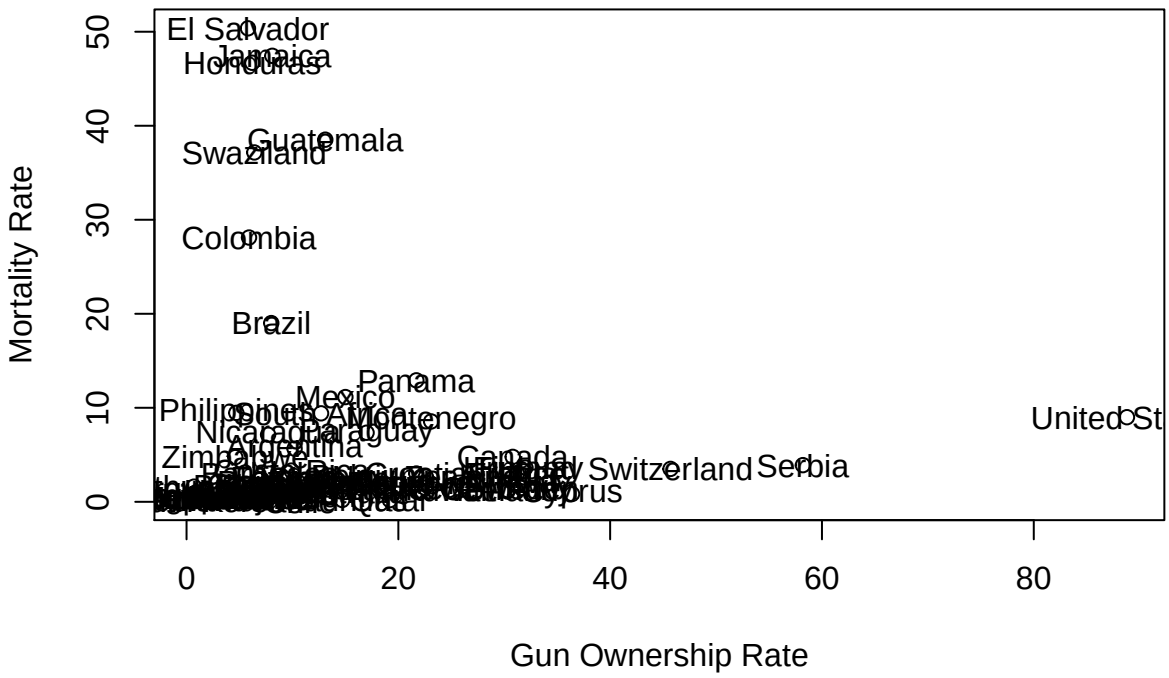


Practice

- Open *gun_violence* dataset and name your object *gun*
- Look at the head and tail of the dataset as well as the summary
- Make a plot to look at the relationship betwen *mortality_rate*, *ownership_rate*

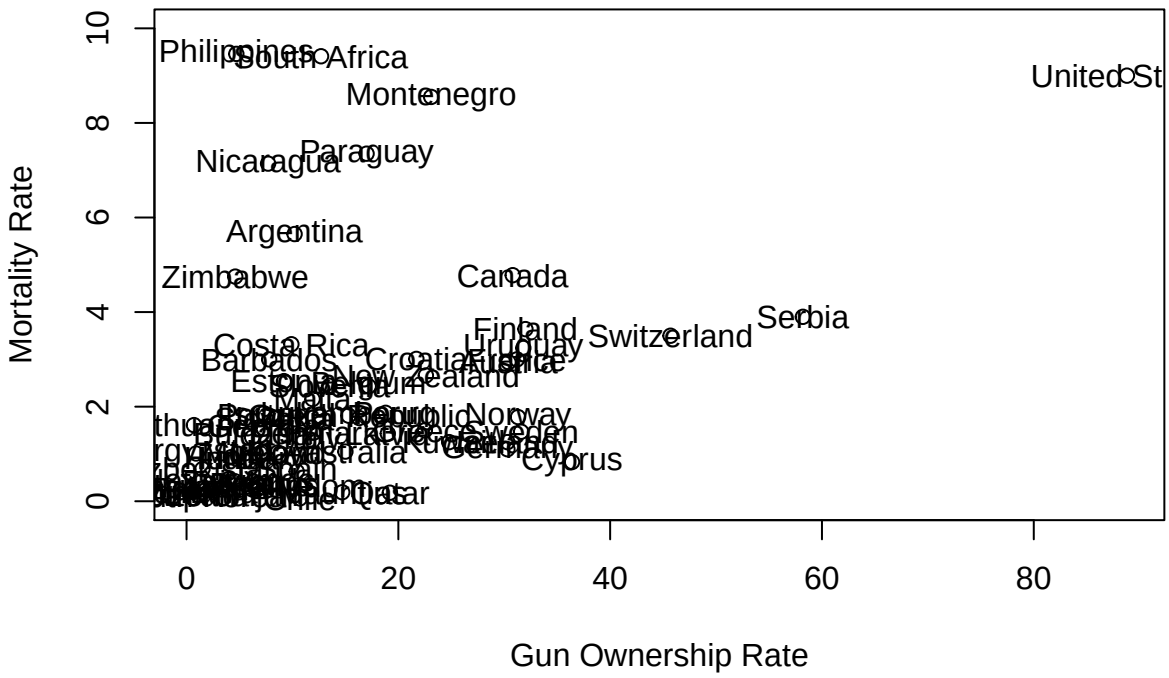
Practice

```
plot(gun$ownership_rate, gun$mortality_rate,
     xlab="Gun Ownership Rate",
     ylab="Mortality Rate")
text(gun$ownership_rate,gun$mortality_rate,
     labels= gun$country)
```



Practice

```
plot(gun$ownership_rate, gun$mortality_rate,
     xlab="Gun Ownership Rate", ylab="Mortality Rate",
     ylim=c(0,10))
text(gun$ownership_rate,gun$mortality_rate,
     labels= gun$country)
```


Mortality Rate
Gun Ownership Rate
0
2
4
6
8
10
0
20
40
60
80
Philippines
South Africa
Montenegro
United St
Nicaragua
Paraguay
Argentina
Zimbabwe
Canada
Serbia
Finland
Switzerland
Costa Rica
Barbados
Croatia
Estonia
Norway
Cyprus