BlainHW1

Blain Morin

February 13, 2018

Question 1: Wine

The data in the file wine.csv (in the datasets folder on Canvas) give the average wine consumption rates (in liters per person) and number of ischemic heart attack deaths (per 1000 men aged 55 to 64 years) for 18 industrialized countries.

Analyze the data and write a brief report that includes a summary of findings, a graphical display and a section describing the methods used to answer the questions of interest.

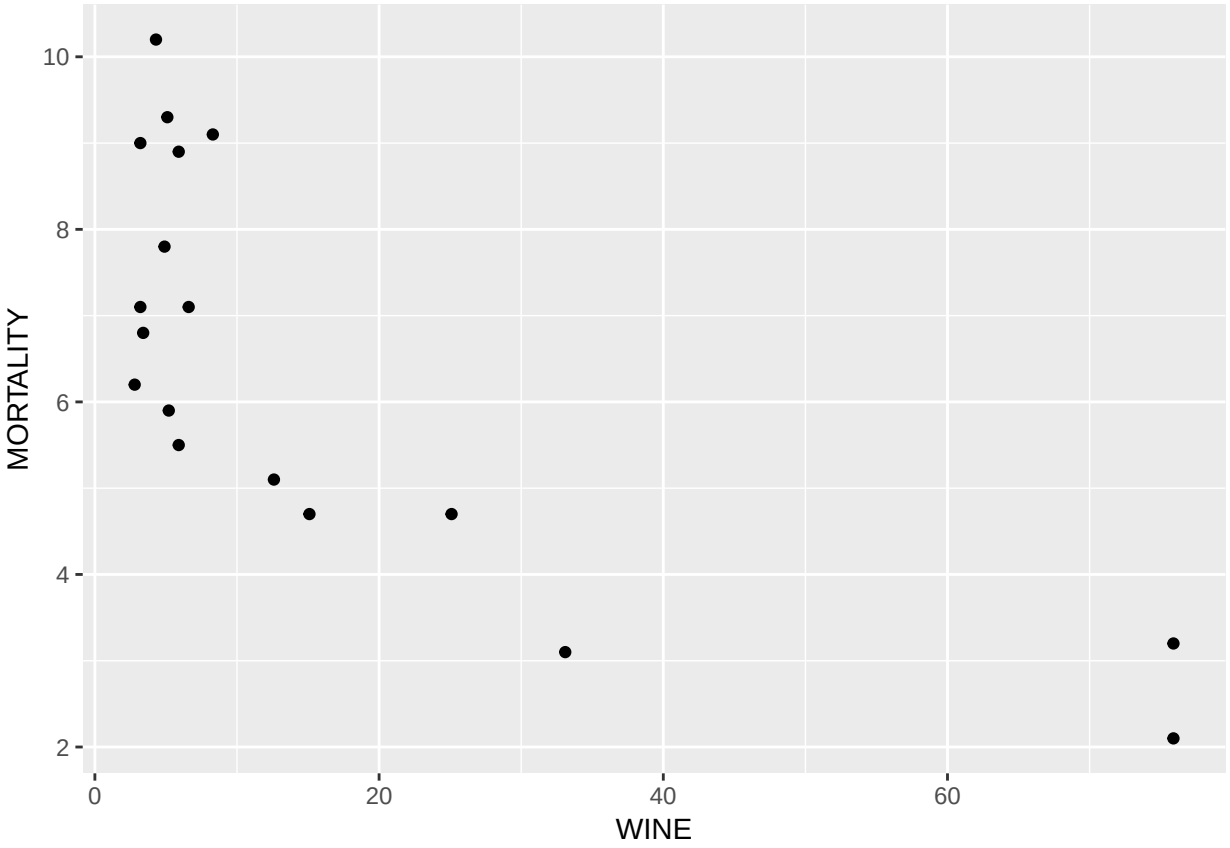
First, lets explore the wine data.

```
## [1] 18  3
```

```
##    COUNTRY               WINE         MORTALITY
##  Length:18          Min.   : 2.80   Min.   : 2.100
##  Class :character   1st Qu.: 4.45   1st Qu.: 4.800
##  Mode  :character   Median : 5.90   Median : 6.500
##                     Mean   :16.47   Mean   : 6.433
##                     3rd Qu.:14.47   3rd Qu.: 8.625
##                     Max.   :75.90   Max.   :10.200
```

We see that the data set contains wine consumption and heart disease mortality rates for 18 different countries. The min, max, median, mean, and qurtiles are shown in the above table.

Next, let's visualize the data using a scatter plot:



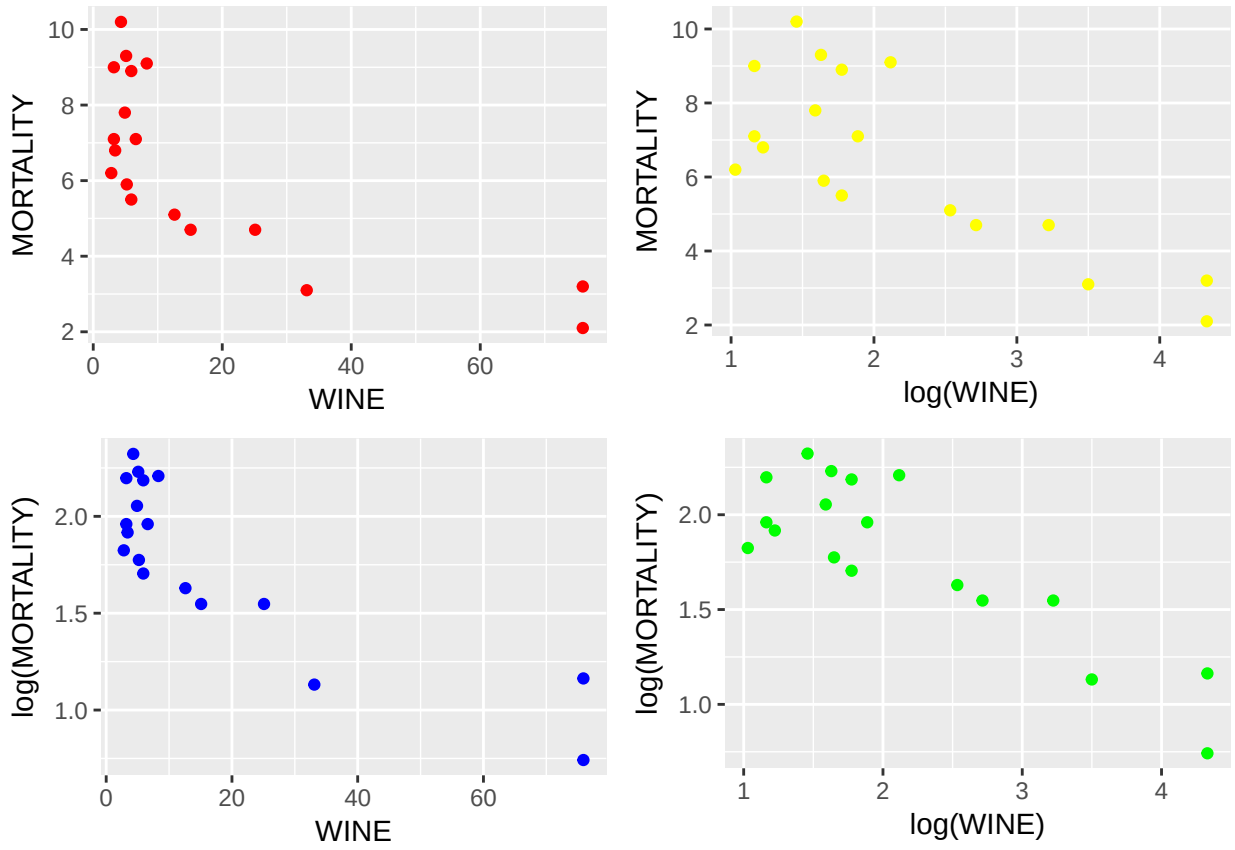
Do these data suggest that heart disease death rates are associated with average wine consumption? If so, how can that be described?

We can see from the graph that there seems to be a negative correlation between wine consumption and heart disease death rates. Computing the correlation confirms the inverse relationship:

```
cor(WINE, MORTALITY)
```

```
## [1] -0.7455682
```

Although there is a strong negative correlation, the relationship does not seem to be linear. We can try to make a stronger linear relationship by using log transforms:



Visually, it seems that using log wine consumption and using both log wine and log mortality have the best linear appearance. Next, lets run a model on the set of transformed data.

```
## Fitted models have different coefficients. Grouping may not work properly. Set `group.pred = FALSE` i
```

MORTALITY

MORTALITY

log(MORTALITY)

log(MORTALITY)

B

CI

p

B

CI

p

B

CI

p

B

CI

p

(Intercept)

7.69

6.68 – 8.69

<.001

10.28

8.52 – 12.04

<.001

2.05

1.90 – 2.19

<.001

2.56

2.29 – 2.82

<.001

WINE

-0.08

-0.11 – -0.04

<.001

-0.02

-0.02 – -0.01

<.001

log(WINE)

-1.77

-2.51 – -1.04

<.001

-0.36

-0.47 – -0.24

<.001

Observations

18

18

18

18

R2 / adj. R2

.556 / .528

.620 / .596

.718 / .700

.738 / .722