APPM2720 Homework 2 solutions

```
library( dataWorkshop)
```

```
## Loading required package: maps
```

```
## Loading required package: fields
```

```
## Loading required package: spam
```

```
## Loading required package: grid
```

```
## Spam version 1.4-0 (2016-08-29) is loaded.
## Type 'help( Spam)' or 'demo( spam)' for a short introduction
## and overview of this package.
## Help for individual functions is also obtained by adding the
## suffix '.spam' to the function name, e.g. 'help( chol.spam)'.
```

```
##
## Attaching package: 'spam'
```

```
## The following objects are masked from 'package:base':
##
##     backsolve, forwardsolve
```

```
data(BoulderTemperature)
data( AudiA4)
```

(1) Solution is self evident!

(2) Load the A4 data set into your R session and create mileage data set: `mileage <- AudiA4[,2]` If you divide the mileage by 1000 how do the mean and median change?

These stats scale by 1000. E.g. `median (mileage/1000)  == median( mileage)/1000`

How does the interquartile range and the standard deviation change?

Also scale by 1000.

If you subtract 1400 from the mileage how does the median change?

`median( mileage - 1400) == median( mileage) - 1400`

How does the interquartile range change?

The IQR does **not** change because shifting the data by a constant does not effect the spread.

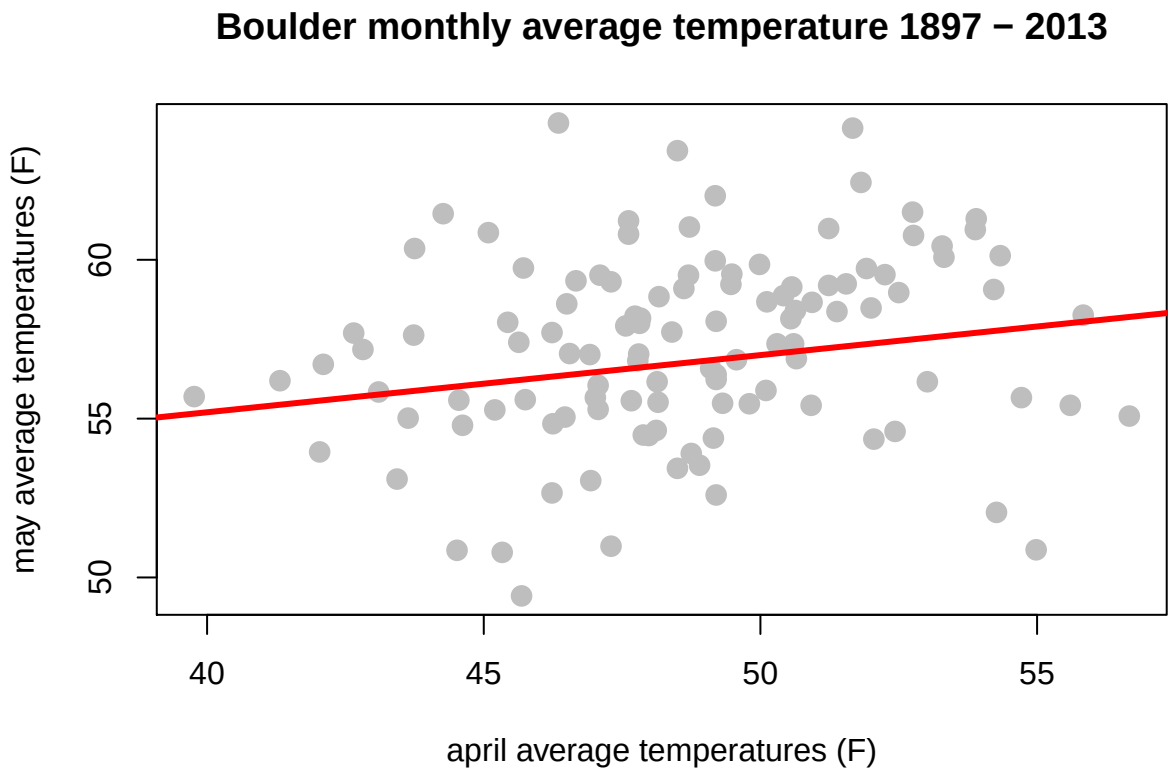
(3) For (`BoulderTemperature.rda`) make a plot of April temperatures against May temperatures.

(4) Use the abline function to add a line through your scatterplot that summarizes the relationship.

I don't see a strong relationship between the months but there may be slight positive dependence. I.e. colder April implies a colder May and a warmer Apr implies a warmer May. But there is quite a bit of varibility. For example one of the warmest April means (55 F) is followed by a cold May (~51 F).

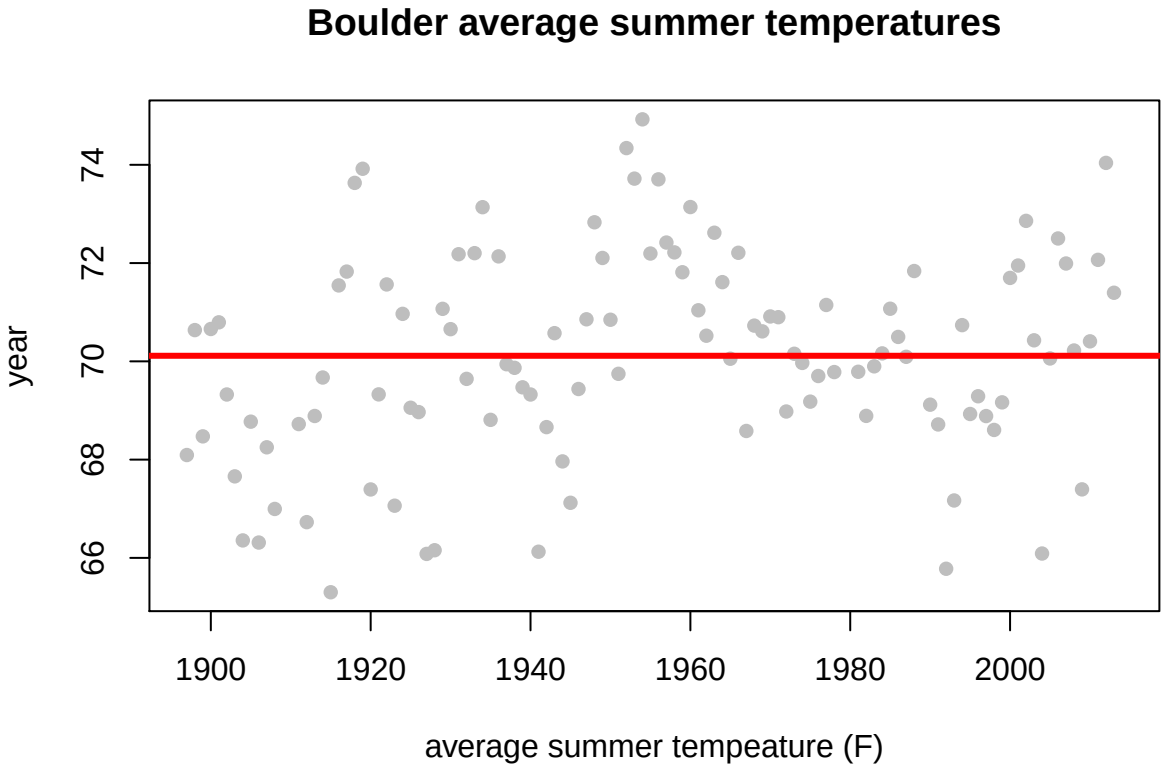
```
plot(BoulderTemperature$apr, BoulderTemperature$may,
ylab="may average temperatures (F)",
xlab= "april average temperatures (F)",
# some nice options to make the plot clearer
pch=16, cex=1.5, col="grey")

title("Boulder monthly average temperature 1897 - 2013")
```

```r
# nice option is to make line thicker
abline( 48, .18, col="red", lwd= 3)
```



(5) Average the june, july, and august temperatures together for each year. Do you see any obvious trend in these summer temperatures over time?

```r
JJA<- (BoulderTemperature[,6]+
      BoulderTemperature[,7]+
      BoulderTemperature[,8]
      )/3
year<- 1897:2014
plot( year, JJA,
      xlab="average summer tempeature (F)",
      ylab="year", col="grey", pch=16
    )
title("Boulder average summer temperatures")
# optional reference line
abline( mean( JJA, na.rm=TRUE), 0,
        col="red", lwd=3)
```



There are no strong patterns although it does seem that the period 1940 through 1960 tended to be warmer.

(6) Write a function that takes a data set and returns the 6 values: min, first quartile, median , mean , third quartile and max.

There are several ways to write this here I took a style that leverages the quantile and range functions. Note default is to not removes NAs

```
myStats<- function( Y, na.rm=FALSE){
Qs<- quantile( Y, c(.25, .50, .75),
                      na.rm = na.rm)
range<- range(Y, na.rm = na.rm )
meanY<- mean( Y, na.rm = na.rm)
out<- c(range[1],
        Qs[1],Qs[2], meanY, Qs[3],
        range[2] )
# A nice option: add names to this vector
names( out)<- c("min",
                "Q1", "mean", "median", "Q3",
                "max")
return( out)
}
```

Test it out.

```
myStats(1:11)
```

```
##    min     Q1   mean median     Q3    max
##    1.0    3.5    6.0    6.0    8.5   11.0
```