First Pair Assignment Markdown, Swiss Data Set

Ben Gaiser & Jeremy Russell

10/6/2016

Set global options for viewability

```
options(max.print = 100)
getOption("max.print")
```

```
## [1] 100
```

Load packages and remove from global environment

```
packages <- c('repmis', 'knitr', 'rmarkdown', 'Hmisc', 'ggplot2', 'pastecs')
for (p in packages) {
  if (p %in% installed.packages()) require(p, character.only=TRUE) # what does that [] mean
  else {
    install.packages(p)
    require(p, character.only=TRUE)
  }
}
```

```
## Loading required package: repmis
```

```
## Loading required package: knitr
```

```
## Loading required package: rmarkdown
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
## Loading required package: pastecs
```

```
## Loading required package: boot
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:survival':
##
##     aml
```

```
## The following object is masked from 'package:lattice':
##
##     melanoma
```

```
repmis::LoadandCite(packages, file = 'RpackageCitations.bib')
```

```
## tweaking Hmisc
```

```
rm(packages, p)
```

Set working directories and remove from global environment

```
wrkdir <- c('C:/Users/Benji/Desktop/Statistics/Git/Repositories/data/alcohol-consumption',
            '~/Hertie School/Fall 2016/CollaborativeSocialScienceDataAnalysis/CSSR')
repmis::set_valid_wd(wrkdir)
```

```
## Working directory set to: ~/Hertie School/Fall 2016/CollaborativeSocialScienceDataAnalysis/CSSR
```

```
rm(wrkdir)
```

Import data frames

```
swiss <- as.data.frame(swiss)
```

Initial Descriptive Stats

```
summary(swiss$Fertility)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   35.00   64.70   70.40   70.14   78.45   92.50
```

```r
summary(swiss$Education)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    6.00    8.00   10.98   12.00   53.00
```

```r
describe(swiss$Fertility)
```

```
## swiss$Fertility
##       n missing  unique    Info    Mean     .05     .10     .25     .50
##      47       0      46       1   70.14   47.58   56.24   64.70   70.40
##     .75     .90     .95
##   78.45   84.60   90.67
##
## lowest : 35.0 42.8 44.7 54.3 55.7, highest: 85.8 87.1 92.2 92.4 92.5
```

```r
describe(swiss$Education)
```

```
## swiss$Education
##       n missing  unique    Info    Mean     .05     .10     .25     .50
##      47       0      19    0.99   10.98     2.0     3.0     6.0     8.0
##     .75     .90     .95
##    12.0    23.2    29.0
##
##           1 2 3 5 6  7 8 9 10 11 12 13 15 19 20 28 29 32 53
## Frequency 1 3 4 2 4  7 4 3  2  1  5  3  1  1  1  1  2  1  1
## %         2 6 9 4 9 15 9 6  4  2 11  6  2  2  2  2  4  2  2
```

```r
var(swiss$Fertility)
```

```
## [1] 156.0425
```

```r
var(swiss$Education)
```

```
## [1] 92.45606
```

```r
sd(swiss$Fertility)
```

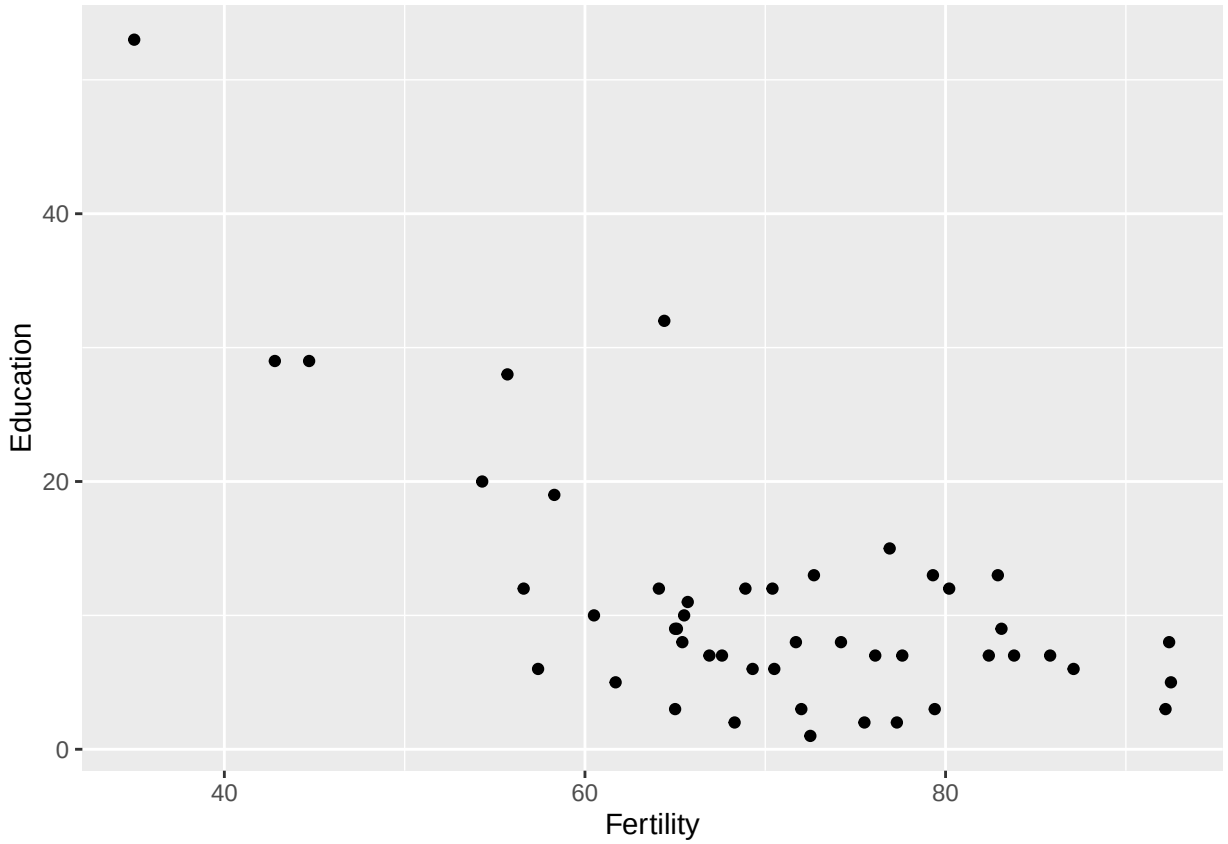
```
## [1] 12.4917
```

```r
sd(swiss$Education)
```

```
## [1] 9.615407
```

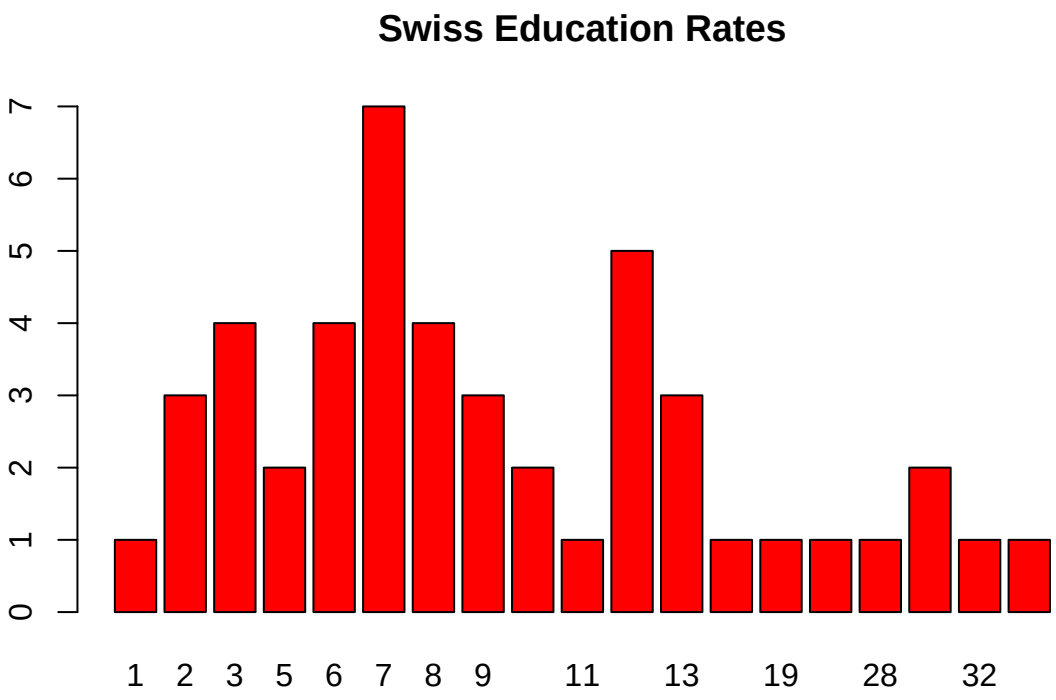
Plot fertility and education with q and barplots. Note: a barplot didn't work with the data for swiss$Fertility. It's basically completely flat.

```
qplot(swiss$Fertility, swiss$Education, xlab = "Fertility",
      ylab = "Education")
```



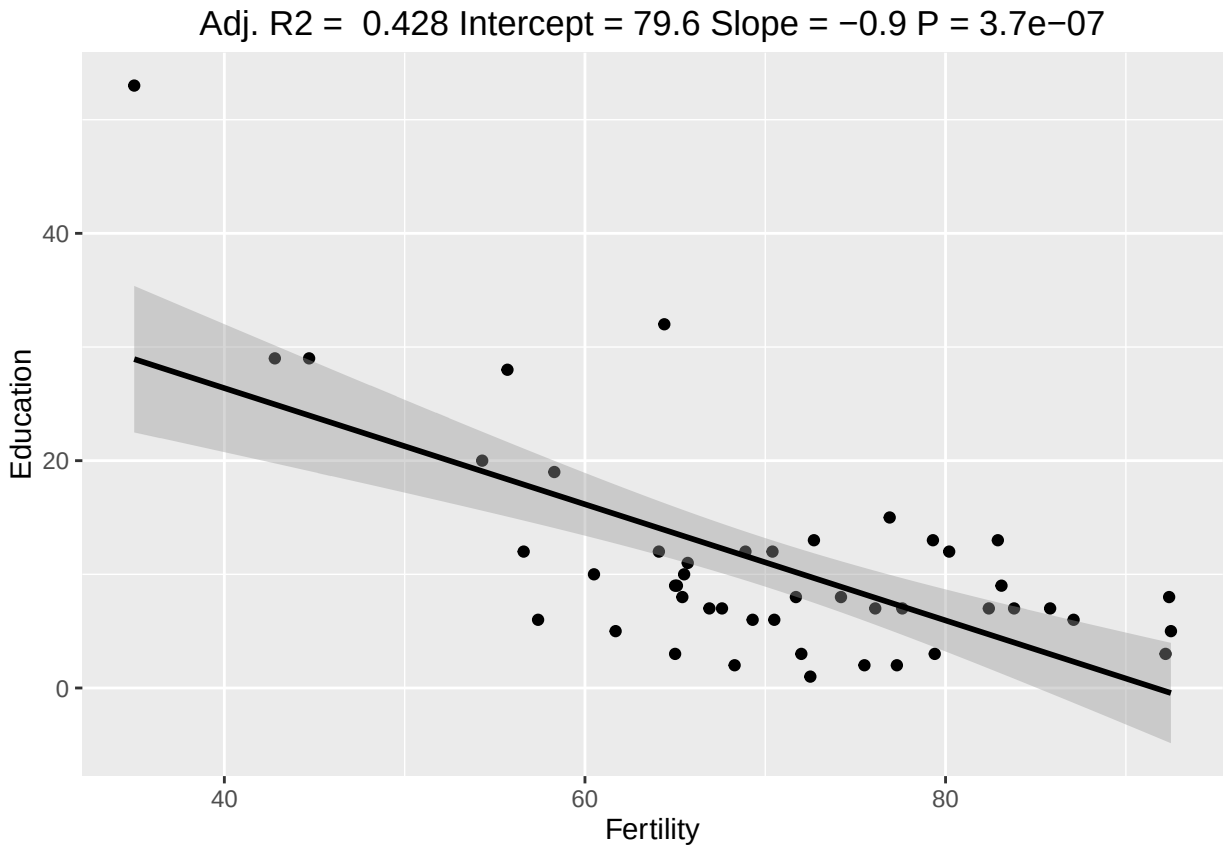
Need to add Swiss canton names to graph below

```
EducationTable <- table(swiss$Education)
barplot(EducationTable, main="Swiss Education Rates",
        xlab="", col=c("red"))
```



Plot fertility and education with ggplot

```
ggplotRegSwiss <- function(fit){
  ggplot(swiss, aes(Fertility, Education)) +
    geom_point() +
    scale_colour_discrete(name="") +
    stat_smooth(method = "lm", col = "black") +
    labs(title = paste("Adj. R2 = ",signif(summary(fit)$adj.r.squared, 3),
                       "Intercept =" ,signif(fit$coef[[1]],3 ),
                       "Slope =" ,signif(fit$coef[[2]], 1),
                       "P =" ,signif(summary(fit)$coef[2,4], 2)))
}
FitOfDataSwiss <- lm(Fertility ~ Education, data = swiss)
ggplotRegSwiss(FitOfDataSwiss)
```


Adj. R2 = 0.428 Intercept = 79.6 Slope = −0.9 P = 3.7e−07
Education
40
20
0
40
60
80
Fertility