Ex-2

Election data

```
### preparing data
data <- read.csv("turnout.csv")
```

1. Check the data

```
dim(data) # check the dimension of the data
```

```
## [1] 14  9
```

```
names(data) # check the variables' name in the data
```

```
## [1] "year"     "VEP"      "VAP"      "total"    "ANES"     "felons"   "noncit"
## [8] "overseas" "osvoters"
```

```
length(data) # the number of columns
```

```
## [1] 9
```

```
print(data$year) # check a specific column by calling the column name
```

```
##  [1] 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2008
```

```
print(max(data$year) - min(data$year))
```

```
## [1] 28
```

```
summary(data) # description of each column
```

```
##       year           VEP              VAP             total
##  Min.   :1980   Min.   :159635   Min.   :164445   Min.   : 64991
##  1st Qu.:1986   1st Qu.:171192   1st Qu.:178930   1st Qu.: 73179
##  Median :1993   Median :181140   Median :193018   Median : 89055
##  Mean   :1993   Mean   :182640   Mean   :194226   Mean   : 89778
##  3rd Qu.:2000   3rd Qu.:193353   3rd Qu.:209296   3rd Qu.:102370
##  Max.   :2008   Max.   :213314   Max.   :230872   Max.   :131304
##
```

```
##       ANES           felons         noncit         overseas       osvoters
##  Min.   :47.00   Min.   : 802   Min.   : 5756   Min.   :1803   Min.   :263
##  1st Qu.:57.00   1st Qu.:1424   1st Qu.: 8592   1st Qu.:2236   1st Qu.:263
##  Median :70.50   Median :2312   Median :11972   Median :2458   Median :263
##  Mean   :65.79   Mean   :2177   Mean   :12229   Mean   :2746   Mean   :263
##  3rd Qu.:73.75   3rd Qu.:3042   3rd Qu.:15910   3rd Qu.:2937   3rd Qu.:263
##  Max.   :78.00   Max.   :3168   Max.   :19392   Max.   :4972   Max.   :263
##                                                                NA's   :13
```

2. Calculating the voting rate

```
# Solution 1: add a new column
data$vap_rates <- 100*(data$total/(data$VAP+data$overseas))
data$vep_rates <- 100*(data$total/(data$VEP))
data[c("year", "vap_rates", "vep_rates")]
```

```
##    year vap_rates vep_rates
## 1  1980  52.03972  54.19551
## 2  1982  40.24522  42.13701
## 3  1984  52.53748  55.24860
## 4  1986  36.07845  38.14115
## 5  1988  49.72260  52.76848
## 6  1990  35.93884  38.41895
## 7  1992  54.04097  58.11384
## 8  1994  38.03086  41.12625
## 9  1996  47.53376  51.65793
## 10 1998  34.83169  38.09316
## 11 2000  49.34211  54.22449
## 12 2002  35.82850  39.51064
## 13 2004  54.54777  60.10084
## 14 2008  55.67409  61.55433
```

```
# Solution 2: assign new variables to a vector
VAPtr <- data$total/(data$VAP + data$overseas) * 100
names(VAPtr) <- data$year
VAPtr
```

```
##     1980     1982     1984     1986     1988     1990     1992     1994
## 52.03972 40.24522 52.53748 36.07845 49.72260 35.93884 54.04097 38.03086
##     1996     1998     2000     2002     2004     2008
## 47.53376 34.83169 49.34211 35.82850 54.54777 55.67409
```

```
VEPtr <- data$total/data$VEP * 100
names(VEPtr) <- data$year
VEPtr
```

```
##     1980     1982     1984     1986     1988     1990     1992     1994
## 54.19551 42.13701 55.24860 38.14115 52.76848 38.41895 58.11384 41.12625
##     1996     1998     2000     2002     2004     2008
## 51.65793 38.09316 54.22449 39.51064 60.10084 61.55433
```

3. Difference between VAP/VEP and ANES

```
# Solution 1
print(summary(data$ANES - data$vap_rates))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.06   18.22   20.62   20.33   22.42   26.17
```

```
print(summary(data$ANES - data$vep_rates))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.581  15.267  16.893  16.836  18.529  22.489
```

```
# Solution 2
# ANES vs. VAP

diffVAP <- data$ANES - VAPtr
mean(diffVAP)
```

```
## [1] 20.32914
```

```
range(diffVAP)
```

```
## [1] 11.06116 26.17150
```

```
# ANES vs. VEP
diffVEP <- data$ANES - VEPtr
mean(diffVEP)
```

```
## [1] 16.83634
```

```
range(diffVEP)
```

```
## [1]  8.581054 22.489359
```

4. Mideterm and president election

```
midterm <- subset(data, year%%4 != 0)
pre <- subset(data, year%%4 == 0)

# Solution 1
diff1 <- mean(midterm$vep_rates) - mean(midterm$ANES)
diff2 <- mean(pre$vep_rates) - mean(pre$ANES)
diff1; diff2
```

```
## [1] -15.4288
```

```
## [1] -17.892
```

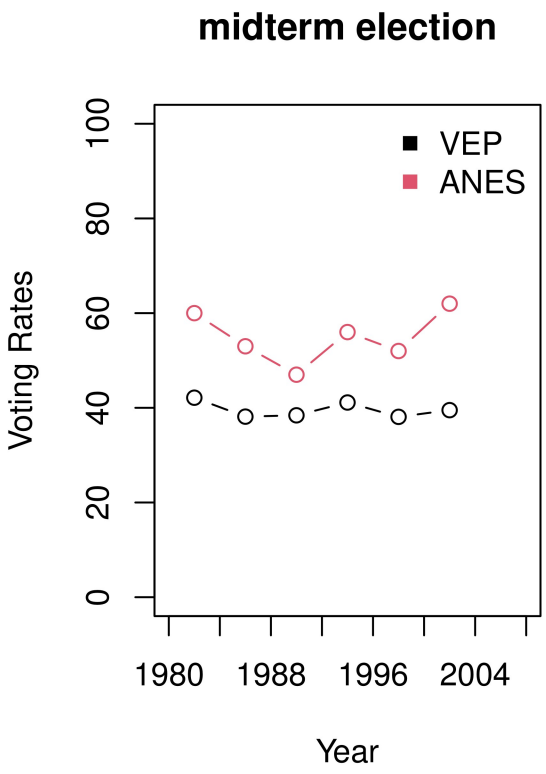
```r
# Solution 2

n.obs <- nrow(data)
## presidential
pdiffVEP <- diffVEP[c(seq(from = 1, to = n.obs, by = 2), 14)]

## midterm (1982 - 2002)
mdiffVEP <- diffVEP[seq(from = 2, to = n.obs - 1, by = 2)]

## comparison
summary(pdiffVEP)
```

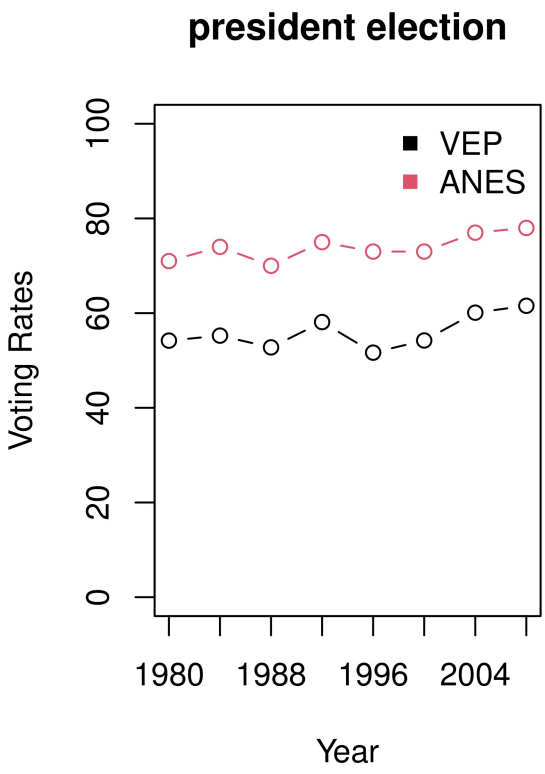
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.45   16.87   17.07   17.89   18.76   21.34
```

```r
summary(mdiffVEP)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.581  14.145  14.866  15.429  17.116  22.489
```

```r
# we can also graph the difference


par(mfrow=c(1,2))

#graphing midterm election

plot(midterm$year, midterm$vep_rates,
     main = "midterm election",
     col=1, type="b",
     xlim=c(1980, 2008), ylim=c(0, 100),
     xlab = "Year", ylab="Voting Rates", xaxt ="n")
axis(side=1, at=seq(from=1980, to=2008, by=4))
lines(midterm$year, midterm$ANES, col=2, type="b")
legend("topright", pch=c(15,15), legend=c("VEP", "ANES"), col=c(1, 2), bty="n")

#graphing president election

plot(pre$year, pre$vep_rates,
     main = "president election",
     col=1, type="b",
     xlim=c(1980, 2008), ylim=c(0, 100),
     xlab = "Year", ylab="Voting Rates", xaxt ="n")
axis(side=1, at=seq(from=1980, to=2008, by=4))
lines(pre$year, pre$ANES, col=2, type="b")
legend("topright", pch=c(15,15), legend=c("VEP", "ANES"), col=c(1, 2), bty="n")
```





5. different period

```
# Solution 1

p_1 <- subset(data, data$year < 1980 + ((max(data$year) - min(data$year))/2))
p_2 <- subset(data, data$year >= 1980 + ((max(data$year) - min(data$year))/2))

diff1 <- mean(p_1$vep_rates) - mean(p_1$ANES)
diff2 <- mean(p_2$vep_rates) - mean(p_2$ANES)

diff1; diff2
```

```
## [1] -15.85378
```

```
## [1] -17.81891
```

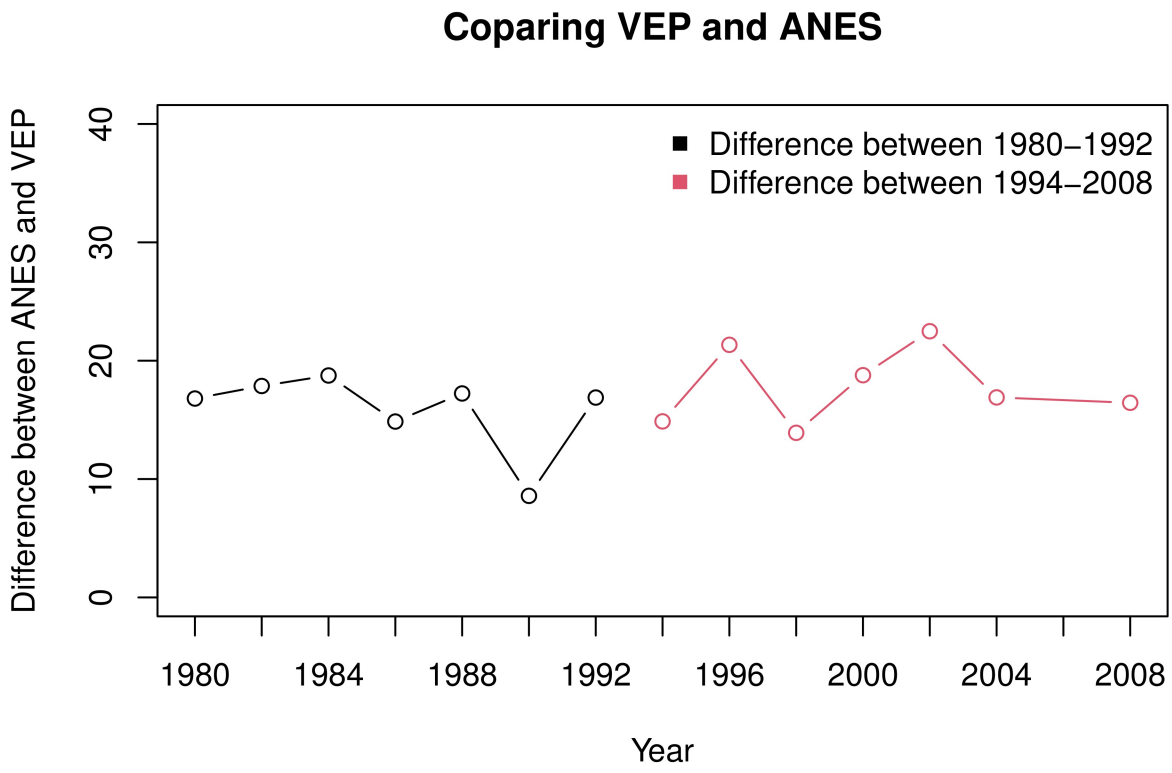
```
# Solution 2

earlier.diffVEP <- diffVEP[1:(n.obs/2)] # first half
later.diffVEP <- diffVEP[(n.obs/2 + 1):n.obs] # second half

## comparison
summary(earlier.diffVEP)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.581  15.832  16.886  15.854  17.547  18.751
```

```
summary(later.diffVEP)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.91   15.66   16.90   17.82   20.06   22.49
```

```
# we can also graph the difference
par(mfrow=c(1,1))

plot(p_1$year, p_1$ANES - p_1$vep_rates, main = "Coparing VEP and ANES",
     col=1, type="b", xlim=c(1980, 2008), ylim=c(0, 40),
     xlab = "Year", ylab="Difference between ANES and VEP", xaxt ="n")
axis(side=1, at=seq(from=1980, to=2008, by=2))
lines(p_2$year, p_2$ANES - p_2$vep_rates, col=2, type="b")
legend("topright", pch=c(15,15),
       legend=c("Difference between 1980-1992", "Difference between 1994-2008"),
       col=c(1, 2), bty="n")
```



6. comparing

```r
# Solution 1

year2008 <- subset(data, data$year == 2008)
adjust_vap <- (year2008$VAP - year2008$felons - year2008$noncit)
year2008$adjust_vap_rates <- 100*((year2008$total - year2008$osvoters)/adjust_vap)

adj.diff <- year2008$ANES - year2008$adjust_vap_rates

diff.vap <- year2008$ANES - year2008$vap_rates
diff.vep <- year2008$ANES - year2008$vep_rates

adj.diff
```

```
## [1] 15.10082
```

```r
diff.vap
```

```
## [1] 22.32591
```

```r
diff.vep
```

```
## [1] 16.44567
```

```r
# Solution 2

adj.VAP <- data$VAP[14] - (data$felons[14] + data$noncit[14])
adj.VAPtr <- (data$total[14] - data$osvoters[14]) / adj.VAP * 100
adj.diff <- data$ANES[14] - adj.VAPtr
## compare adjusted difference to non-adjusted VAP as well as VEP
adj.diff
```

```
## [1] 15.10082
```

```r
diffVAP[14]
```

```
##     2008
## 22.32591
```

```r
diffVEP[14]
```

```
##     2008
## 16.44567
```

Population

```r
# tidyr and dplyr are concise and efficient way to organize the codes

library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
#Preparing data
kenya <- read.csv("Kenya.csv")
sweden <- read.csv("Sweden.csv")
world <- read.csv("World.csv")
```

Q1. Calculating CBR

```r
# Solution 1

sweden$py.total <- sweden$py.men + sweden$py.women
kenya$py.total <- kenya$py.men + kenya$py.women
world$py.total <- world$py.men + world$py.women

## calculate the CBR
sweden.CBR <- c(sum(sweden$births[1:15]) / sum(sweden$py.total[1:15]),
sum(sweden$births[16:30]) / sum(sweden$py.total[16:30]))
kenya.CBR <- c(sum(kenya$births[1:15]) / sum(kenya$py.total[1:15]),
sum(kenya$births[16:30]) / sum(kenya$py.total[16:30]))
world.CBR <- c(sum(world$births[1:15]) / sum(world$py.total[1:15]),
sum(world$births[16:30]) / sum(world$py.total[16:30]))

## assign labels
names(sweden.CBR) <- c("1950-1955", "2005-2010")
names(kenya.CBR) <- c("1950-1955", "2005-2010")
names(world.CBR) <- c("1950-1955", "2005-2010")

## alternative, more efficient way to assign the same labels
names(sweden.CBR) <- names(kenya.CBR) <- names(world.CBR) <-
c("1950-1955", "2005-2010")

## display results
sweden.CBR
```

```
##  1950-1955  2005-2010
## 0.01539614 0.01192554
```

```
kenya.CBR
```

```
##  1950-1955  2005-2010
## 0.05209490 0.03851507
```

```
world.CBR
```

```
##  1950-1955  2005-2010
## 0.03732863 0.02021593
```

```
# Solution 2

# create a function for repeating calculation
cbr <- function(x, p){
  x<-filter(x, period == p)
  CBR <- sum(x$births)/sum(x$py.men+x$py.women)
  return(round(CBR,4))
}

kenya_cbr <- c(cbr(kenya, "1950-1955"), cbr(kenya, "2005-2010"))
sweden_cbr <- c(cbr(sweden, "1950-1955"), cbr(sweden, "2005-2010"))
world_cbr <- c(cbr(world, "1950-1955"), cbr(world, "2005-2010"))

options(digits = 3) # Retain 3 decimal places
cat('     ', paste("1950-1955", "2005-2010"), '\n',
  ' Kenya:', paste(kenya_cbr, sep=' '), '\n',
    'Sweden:', paste(sweden_cbr, sep=' '), '\n',
    'World:', paste(world_cbr, sep=' '))
```

```
##       1950-1955 2005-2010
##   Kenya: 0.0521 0.0385
##  Sweden: 0.0154 0.0119
##  World: 0.0373 0.0202
```

Q2. Calculating ASFR

```
asfr <- function(x, p){
  x<-filter(x, period == p)
  ASFR <- x$births/x$py.women
  return(ASFR)

}

kenya$asfr <- c(asfr(kenya, "1950-1955"), asfr(kenya, "2005-2010"))
sweden$asfr <- c(asfr(sweden, "1950-1955"), asfr(sweden, "2005-2010"))
world$asfr <- c(asfr(world, "1950-1955"), asfr(world, "2005-2010"))

kenya[c("age", "asfr")][c(4:10, 19:25), ]
```

```
##      age   asfr
## 4  15-19 0.1688
## 5  20-24 0.3560
## 6  25-29 0.3466
## 7  30-34 0.2895
## 8  35-39 0.2064
## 9  40-44 0.1119
## 10 45-49 0.0391
## 19 15-19 0.1006
## 20 20-24 0.2358
## 21 25-29 0.2329
## 22 30-34 0.1809
## 23 35-39 0.1313
## 24 40-44 0.0563
## 25 45-49 0.0382
```

```
sweden[c("age", "asfr")][c(4:10, 19:25), ]
```

```
##      age     asfr
## 4  15-19 0.038909
## 5  20-24 0.127711
## 6  25-29 0.125244
## 7  30-34 0.087364
## 8  35-39 0.048604
## 9  40-44 0.016210
## 10 45-49 0.001342
## 19 15-19 0.005971
## 20 20-24 0.050732
## 21 25-29 0.116209
## 22 30-34 0.132274
## 23 35-39 0.062592
## 24 40-44 0.012160
## 25 45-49 0.000614
```

```
world[c("age", "asfr")][c(4:10, 19:25), ]
```

```
##      age    asfr
## 4  15-19 0.09030
## 5  20-24 0.23763
## 6  25-29 0.25245
## 7  30-34 0.20416
## 8  35-39 0.13811
## 9  40-44 0.06361
## 10 45-49 0.01519
## 19 15-19 0.04849
## 20 20-24 0.15197
## 21 25-29 0.14698
## 22 30-34 0.09381
## 23 35-39 0.04669
## 24 40-44 0.01627
## 25 45-49 0.00451
```

3. Calculating TFR

```
tfr <- function(x, p){
  x <- filter(x, period == p)
  return(sum(x$asfr*5))

}

kenya_tfr <- c(tfr(kenya, "1950-1955"), tfr(kenya, "2005-2010"))
sweden_tfr <- c(tfr(sweden, "1950-1955"), tfr(sweden, "2005-2010"))
world_tfr <- c(tfr(world, "1950-1955"), tfr(world, "2005-2010"))

#changes in women amount
print(sum(world[world$period == "2005-2010", "py.women"])
      - sum(world[world$period == "1950-1955", "py.women"]))
```

```
## [1] 1e+07
```

```
#changes in total births
print(sum(world[world$period == "2005-2010", "births"])
      - sum(world[world$period == "1950-1955", "births"]))
```

```
## [1] 185690
```

```
names(kenya_tfr) <- names(sweden_tfr) <- names(world_tfr) <-
  c("1950-1955", "2005-2010")


kenya_tfr; sweden_tfr; world_tfr
```

```
## 1950-1955 2005-2010
##      7.59      4.88

## 1950-1955 2005-2010
##      2.23      1.90

## 1950-1955 2005-2010
##      5.01      2.54
```

```
sum(world[world$period == "2005-2010", "births"])
```

```
## [1] 674581
```

```
sum(world[world$period == "1950-1955", "births"])
```

```
## [1] 488892
```

4. Calculating CDR

```
cdr <- function(x, p){
  x<-filter(x, period == p)
  CDR <- sum(x$death)/sum(x$py.men+x$py.women)
  return(CDR)
}

kenya_cdr <- c(cdr(kenya, "1950-1955"), cdr(kenya, "2005-2010"))
sweden_cdr <- c(cdr(sweden, "1950-1955"), cdr(sweden, "2005-2010"))
world_cdr <- c(cdr(world, "1950-1955"), cdr(world, "2005-2010"))

names(kenya_cdr) <- names(sweden_cdr) <- names(world_cdr) <-
  c("1950-1955", "2005-2010")

kenya_cdr; sweden_cdr; world_cdr
```

```
## 1950-1955 2005-2010
##    0.0240    0.0104

## 1950-1955 2005-2010
##   0.00984   0.00997

## 1950-1955 2005-2010
##   0.01932   0.00817
```

5. Calculating ASDR

```
asdr <- function(x, p){
  x<-filter(x, period == p)
  ASDR <- x$deaths/(x$py.women + x$py.men)
  return(ASDR)

}

kenya_asdr <- asdr(kenya, "2005-2010")
sweden_asdr <- asdr(sweden, "2005-2010")

summary(kenya_asdr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0029  0.0052  0.0113  0.0240  0.0174  0.1586
```

```
summary(sweden_asdr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0001  0.0005  0.0007  0.0108  0.0038  0.1099
```

```
# if you want to see each cohort
names(kenya_asdr) <- kenya$age[16:30] # period 2005-2010
kenya_asdr
```

```
##     0-4     5-9   10-14   15-19   20-24   25-29   30-34   35-39   40-44   45-49
## 0.02092 0.00291 0.00292 0.00294 0.00389 0.00656 0.01060 0.01388 0.01347 0.01129
##   50-54   55-59   60-69   70-79     80+
## 0.01115 0.01390 0.02540 0.06126 0.15862
```

6. counterfactual CDR

```
kenya$p_counter <- (sweden$py.men+sweden$py.women)/
  sum(sweden[sweden$period == "1950-1955", "py.men"],
      sweden[sweden$period == "1950-1955", "py.women"])

kenya$p_counter[c(16:30)] <- (sweden[sweden$period == "2005-2010", "py.men"] +
                                sweden[sweden$period == "2005-2010", "py.women"])/
  sum(sweden[sweden$period == "2005-2010", "py.men"],
      sweden[sweden$period == "2005-2010", "py.women"])

kenya_cdr_counter <- sum(kenya_asdr*kenya[kenya$period == "2005-2010", "p_counter"])

# original data
kenya_cdr
```

```
## 1950-1955 2005-2010
##    0.0240    0.0104
```

```
# counter facts
kenya_cdr_counter
```

```
## [1] 0.0232
```