```
survey<-read.csv("E:/Repos/StatisticsR/DSC520-Statistics/week4/acs-14-1yr-s0201.csv")

head(survey)


install.packages("dplyr")

library(dplyr)
```

Using the dplyr package, use the 6 different operations to analyze/transform the data - GroupBy, Summarize, Mutate, Filter, Select, and Arrange – Remember this isn't just modifying data, you are learning about your data also – so play around and start to understand your dataset in more detail

```
#select

select(survey, Id, Geography, HSDegree)


#filter

filter(survey, HSDegree>90)

filter(survey, HSDegree>93)


#mutate

mutate(survey,RacesReportedIn1000=RacesReported/1000)


#summarize

survey %>% summarize(mean_HSDegree = mean(HSDegree))


#GroupBy

survey %>% group_by(RacesReported, Geography)


#Arrange

survey %>%  summarize(mean_HSDegree = mean(HSDegree),min_weight = min(HSDegree)) %>%arrange(desc(mean_HSDegree))
```

Using the purrr package – perform 2 functions on your dataset. You could use zip_n, keep, discard, compact, etc.

```
library(purrr)


#keep

survey %>%  map(sample, 5) %>%  keep(~mean(survey$HSDegree) > 80)


#discard

survey %>%  map(sample, 5) %>%  discard(~mean(survey$BachDegree) < 35)
```

Use the cbind and rbind function on your dataset

```
#cbind

cb<-cbind(survey$Geography,survey$HSDegree,survey$BachDegree)

cb


#rbind

rb<-
rbind(survey[1,],survey[2,],survey[3,],survey[4,],survey[5,],survey[6,],su
rvey[7,])

rb


#Split a string, then concatenate the results back together

library(stringr)


#Split

splitstring <- strsplit(as.character(survey$Geography), ",", fixed =
FALSE)

splitstring
```

```
#Bind
pastestring<-paste(splitstring)
pastestring
```