

## Assignment 3

Use the `sc_debt.Rds` dataset to answer the following questions. As always, your assignment must be turned in as a `.Rmd` file.

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.0.5

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.5

## Warning: package 'tibble' was built under R version 4.0.5

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'readr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## Warning: package 'forcats' was built under R version 4.0.5

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

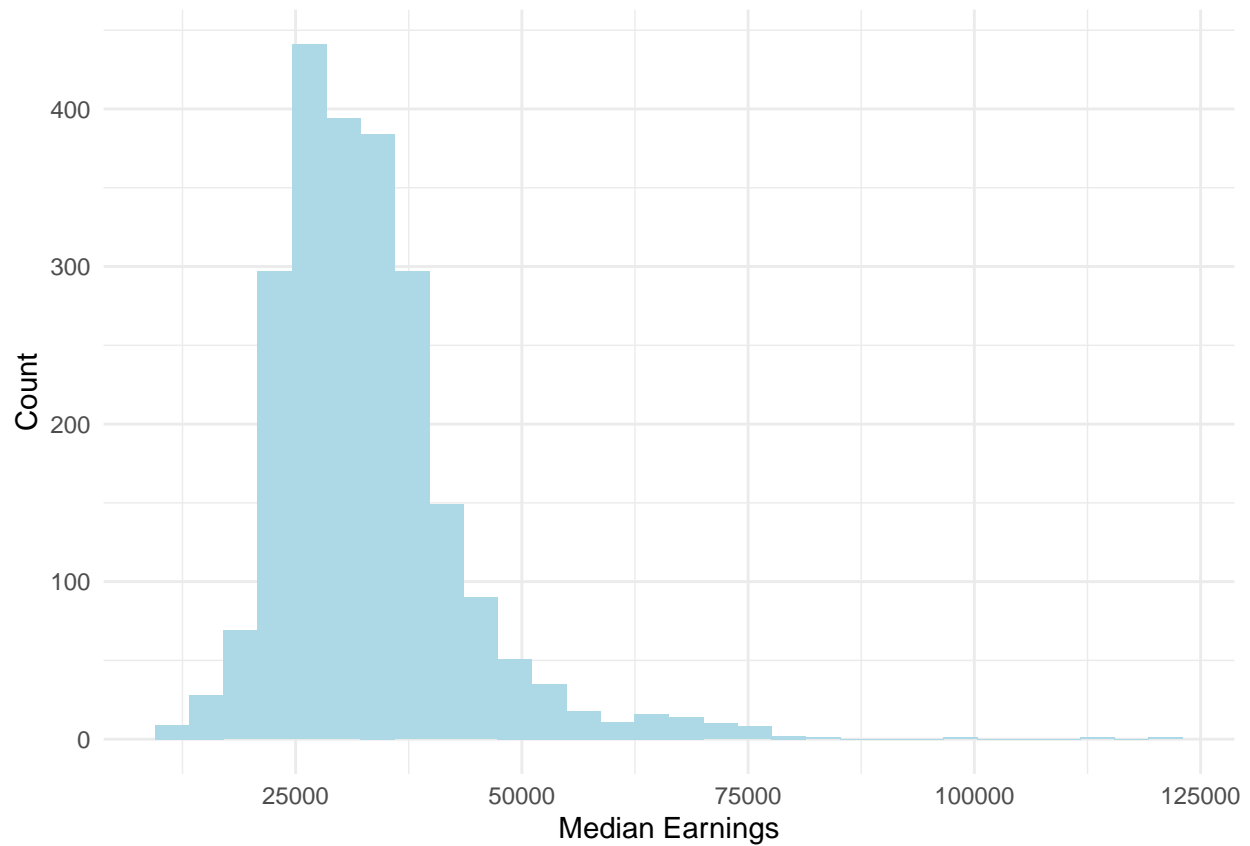
df<-readRDS("sc_debt.Rds")
```

1. Create a graph that shows the distribution of grads' income `md_earn_wne_p6`

```
## Option 1: histogram
df%>%
  ggplot(aes(x=md_earn_wne_p6))+
  geom_histogram(fill="lightblue")+
  xlab("Median Earnings")+ylab("Count")+
  theme_minimal()
```

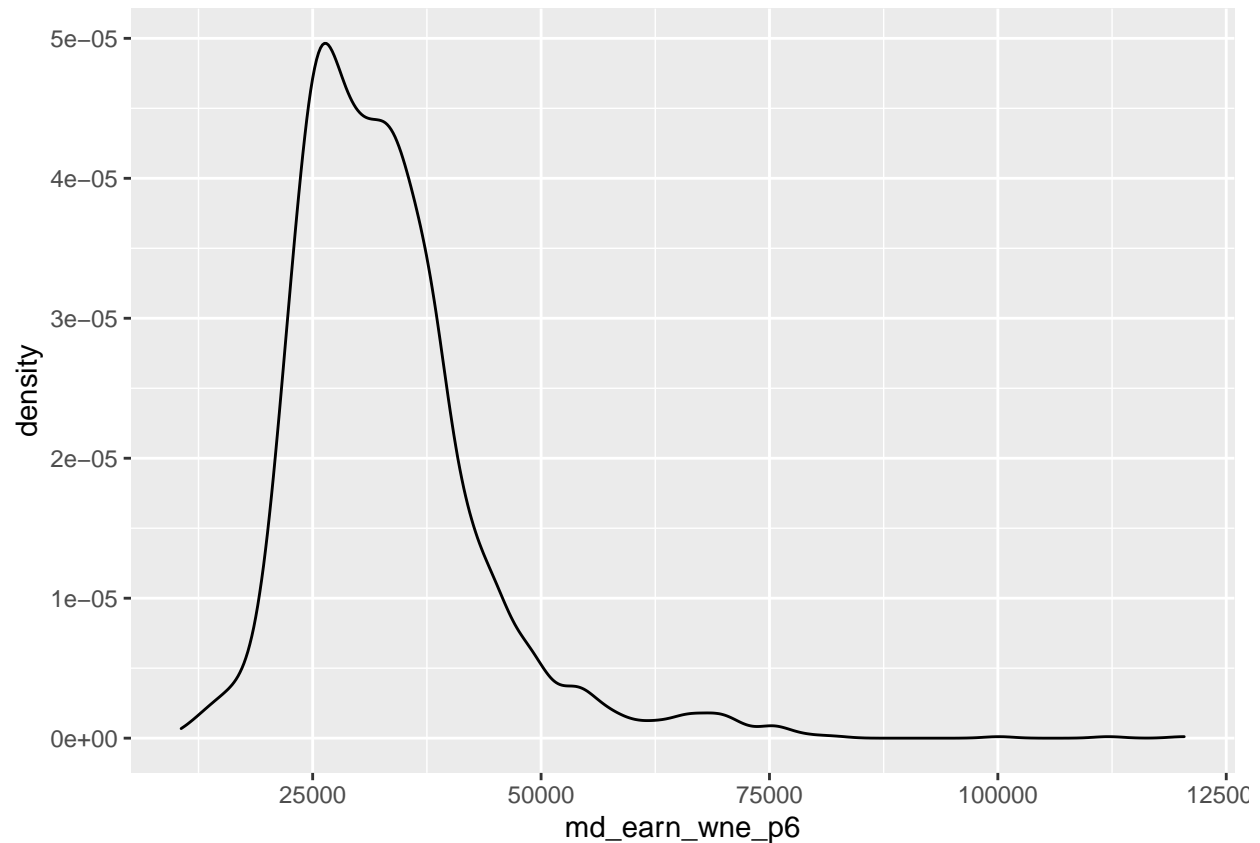
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 228 rows containing non-finite values (stat_bin).
```



```
## Option 2: density plot  
df%>%  
  ggplot(aes(x=md_earn_wne_p6))+  
  geom_density()
```

```
## Warning: Removed 228 rows containing non-finite values (stat_density).
```

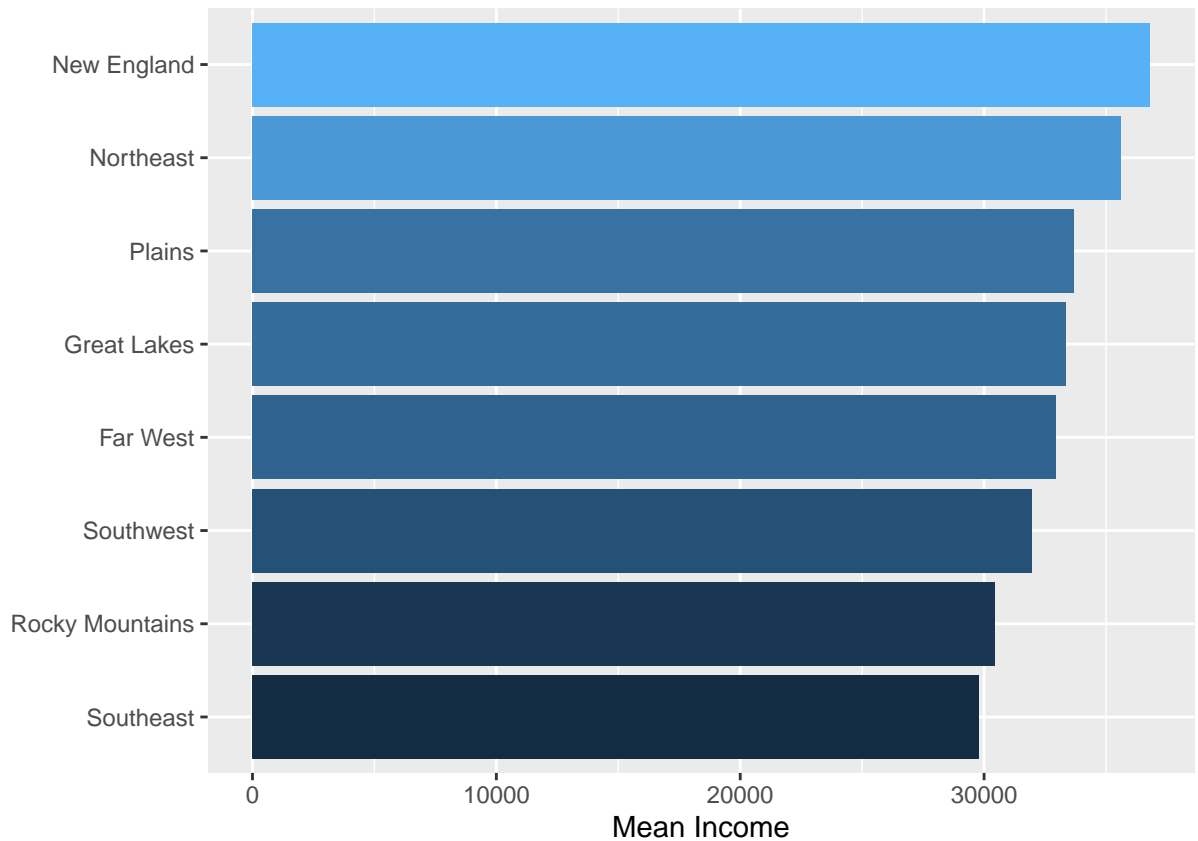


```
df<-df%>%
  mutate(region=ifelse(region=="Southwest","Southeast",region))
```

2. Create a graph that shows the average level of income by region.

```
df_sum<-df%>%
  group_by(region)%>%
  summarize(mean_income=mean(md_earn_wne_p6,na.rm=TRUE))

df_sum%>%
  ggplot(aes(x=fct_reorder(region,mean_income),y=mean_income,fill=mean_income))+
  geom_bar(stat="identity")+
  xlab("")+ylab("Mean Income")+
  coord_flip()+
  theme(legend.position = "none")
```

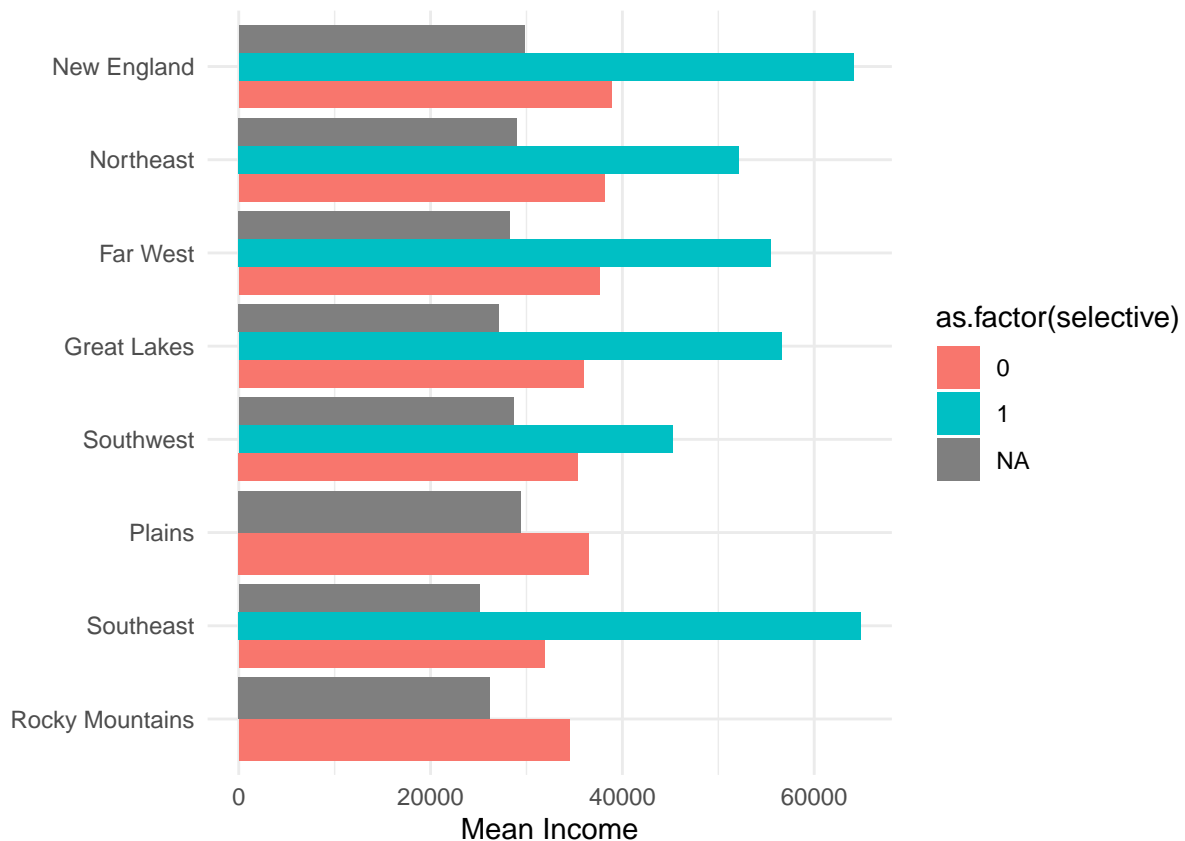


3. Create another graph that shows average level of income by region and selectivity.

```
df_sum<-df%>%
# mutate(selective=ifelse(selective==1,"Selective","Not Selective"))%>%
# mutate(selective=ifelse(is.na(selective),"Not Selective",selective))%>%
group_by(region,selective)%>%
summarize(mean_income=mean(md_earn_wne_p6,na.rm=TRUE))
```

## 'summarise()' has grouped output by 'region'. You can override using the '.groups' argument.

```
df_sum%>%
ggplot(aes(x=fct_reorder(region,mean_income),y=mean_income,fill=as.factor(selective)))+
geom_bar(stat="identity",position="dodge")+
xlab("")+ylab("Mean Income")+
coord_flip()+
theme(legend.position = "none")+
theme_minimal()
```

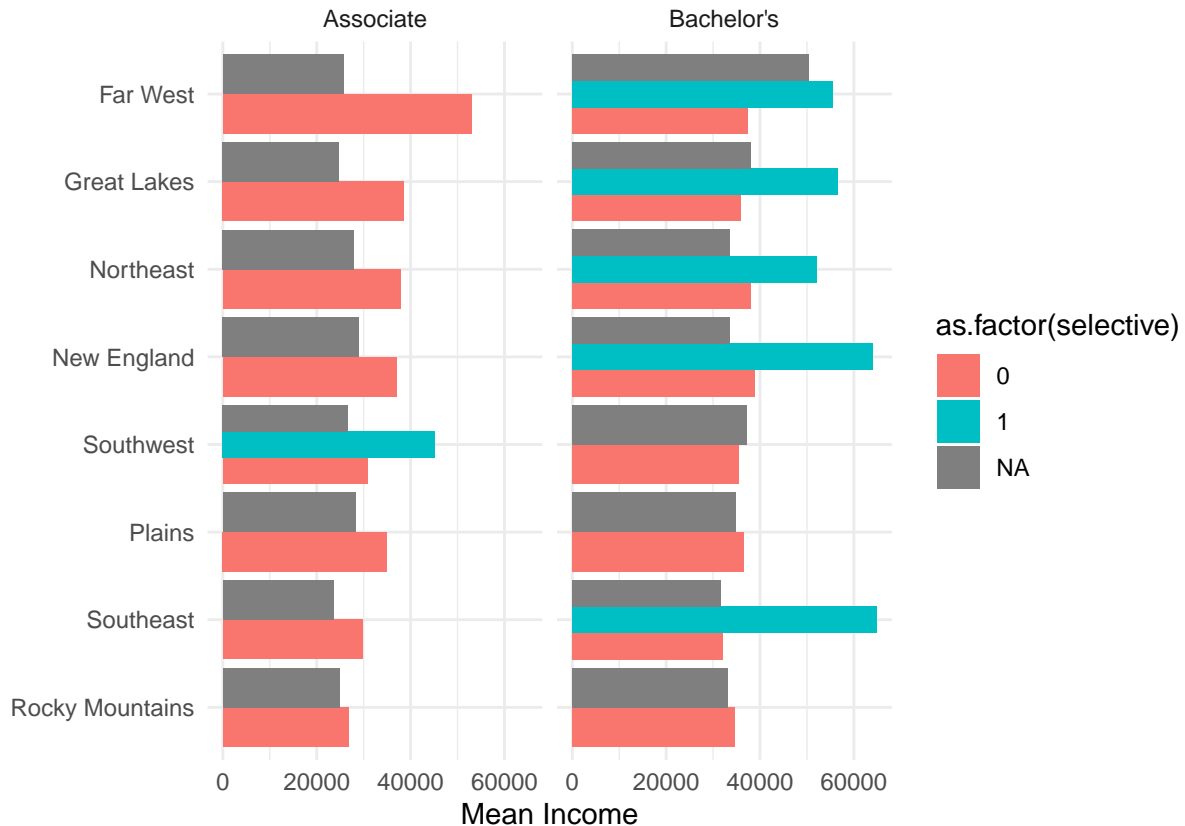


4. Create a graph that shows average levels of income by region, selectivity, and predominant degree.

```
df_sum<-df%>%
# mutate(selective=ifelse(selective==1,"Selective","Not Selective"))%>%
# mutate(selective=ifelse(is.na(selective),"Not Selective",selective))%>%
group_by(region,selective,preddeg)%>%
summarize(mean_income=mean(md_earn_wne_p6,na.rm=TRUE))
```

## 'summarise()' has grouped output by 'region', 'selective'. You can override using the '.groups' argument

```
df_sum%>%
ggplot(aes(x=fct_reorder(region,mean_income),y=mean_income,fill=as.factor(selective)))+
geom_bar(stat="identity",position="dodge")+
xlab("")+ylab("Mean Income")+
coord_flip()+
theme(legend.position = "none")+
facet_wrap(~preddeg)+
theme_minimal()
```



5. Create a graph that shows average levels of income by region, selectivity, and predominant degree and control.

```
df_sum<-df%>%
# mutate(selective=ifelse(selective==1,"Selective","Not Selective"))%>%
# mutate(selective=ifelse(is.na(selective),"Not Selective",selective))%>%
group_by(region,selective,preddeg,control)%>%
summarize(mean_income=mean(md_earn_wne_p6,na.rm=TRUE))
```

## 'summarise()' has grouped output by 'region', 'selective', 'preddeg'. You can override using the '.g

```
df_sum%>%
ggplot(aes(x=fct_reorder(region,mean_income),y=mean_income,fill=as.factor(selective)))+
geom_bar(stat="identity",position="dodge")+
xlab("")+ylab("Mean Income")+
coord_flip()+
theme(legend.position = "none")+
facet_wrap(control~preddeg)+
theme_minimal()
```

## Warning: Removed 1 rows containing missing values (geom\_bar).

