Salary Analysis

Ashen Shavinda

1/14/2022

library(tidyverse)

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

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.1.0 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

Salary <- read.csv("salary.csv")  
View(Salary)

toalNum <- Salary %>% summarise(count=n())  
toalNum

## count  
## 1 3000

summary(Salary)

## age maritl race education   
## Min. :18.00 Length:3000 Length:3000 Length:3000   
## 1st Qu.:33.75 Class :character Class :character Class :character   
## Median :42.00 Mode :character Mode :character Mode :character   
## Mean :42.41   
## 3rd Qu.:51.00   
## Max. :80.00   
## jobclass health health\_ins salary   
## Length:3000 Length:3000 Length:3000 Min. : 20.09   
## Class :character Class :character Class :character 1st Qu.: 85.38   
## Mode :character Mode :character Mode :character Median :104.92   
## Mean :111.70   
## 3rd Qu.:128.68   
## Max. :318.34

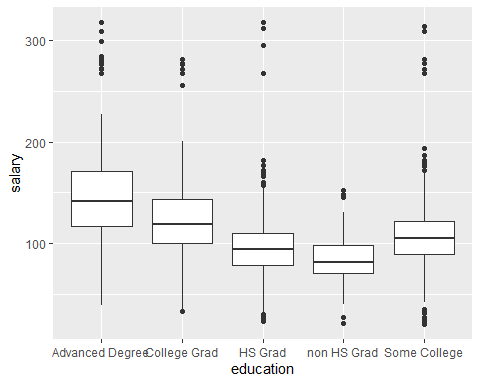
avg\_1 <- Salary %>% summarise(avg\_salary=mean(salary))  
avg\_1

## avg\_salary  
## 1 111.7036

avg\_2 <- Salary %>% group\_by(education) %>% summarise(avg\_salary=mean(salary))  
avg\_2

## # A tibble: 5 x 2  
## education avg\_salary  
## <chr> <dbl>  
## 1 Advanced Degree 151.   
## 2 College Grad 124.   
## 3 HS Grad 95.8  
## 4 non HS Grad 84.1  
## 5 Some College 108.

box\_plot <- ggplot(Salary, aes(x = education, y = salary))  
box\_plot + geom\_boxplot()



pers <- Salary %>% select(health,health\_ins) %>% filter(health == "Good" & health\_ins == "Yes") %>% summarise(percentage = 100\*n()/nrow(Salary))   
pers

## percentage  
## 1 51.16667

counts <- table(Salary$maritl, Salary$jobclass)  
barplot(counts, main="Job classes across the different Martial status",  
 xlab="Job Class",ylab="Martial Status", col=c("darkblue","Gray","Orange","Yellow","Green"),  
 legend = rownames(counts), beside=TRUE)

