DSC520_Week3.1_Guruprasad_Velikadu_Assignment03

Guruprasad Velikadu Krishnamoorthy

2022-12-18

Contents

1 Assignment **1**

1.1 Load the ggplot2 package 1

1.2 Set the working directory to the root of your DSC 520 directory 2

1.3 Load the `data/r4ds/heights.csv` to 2

1.4 https://ggplot2.tidyverse.org/reference/geom_point.html 2

1.5 `age` vs. `earn` 3

1.6 `ed` vs. `earn` 3

1.7 Re-create the three scatterplots and add a regression trend line using 4

1.8 `age` vs. `earn` 5

1.9 `ed` vs. `earn` 6

1.10 Create a scatterplot of `height`` vs.earn. Usesexas thecol‘ (color) attribute 7

1.11 Using `ggtitle()`, `xlab()`, and `ylab()` to add a title, x label, and y label to the previous plot 8

1.12 https://ggplot2.tidyverse.org/reference/geom_histogram.html 9

1.13 Create a histogram of the `earn` variable using `geom_histogram()` 10

1.14 https://ggplot2.tidyverse.org/reference/geom_density.html 11

2 Session Info **12**

1 Assignment

1.1 Load the ggplot2 package

```
library(ggplot2)
theme_set(theme_minimal())
library(formatR)
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
```

1.2 Set the working directory to the root of your DSC 520 directory

```
knitr::opts_knit$set(root.dir = "C:/Users/Gurup/GURU/Learning/Masters/Term_2/DSC520_T302_Statistics_for_
```

1.3 Load the data/r4ds/heights.csv to

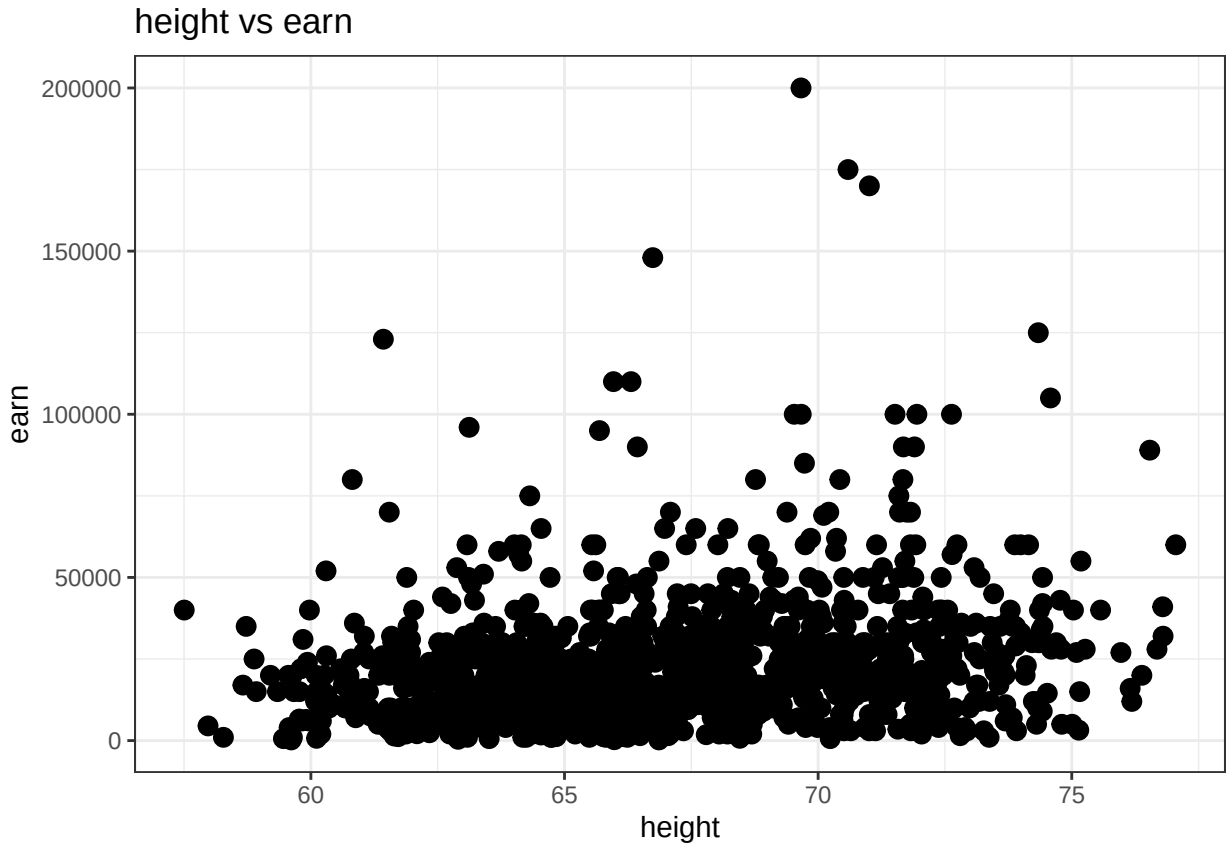
```
heights_df <- read.csv("data/r4ds/heights.csv")
```

1.4 https://ggplot2.tidyverse.org/reference/geom_point.html

1.4.1 Using geom_point() create three scatterplots for

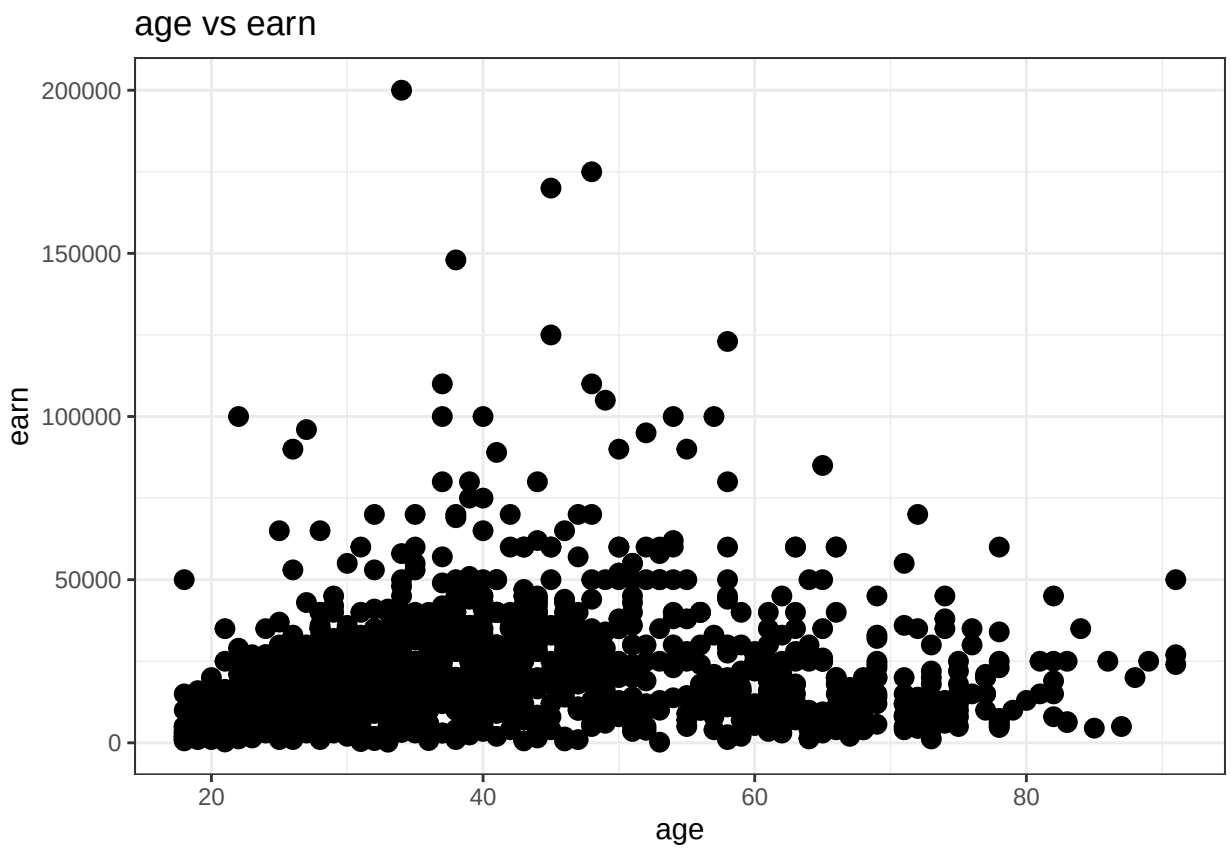
1.4.2 height vs. earn

```
ggplot(data = heights_df, aes(x = height, y = earn)) + geom_point(size = 3) +
    theme_bw() + labs(title = "height vs earn")
```



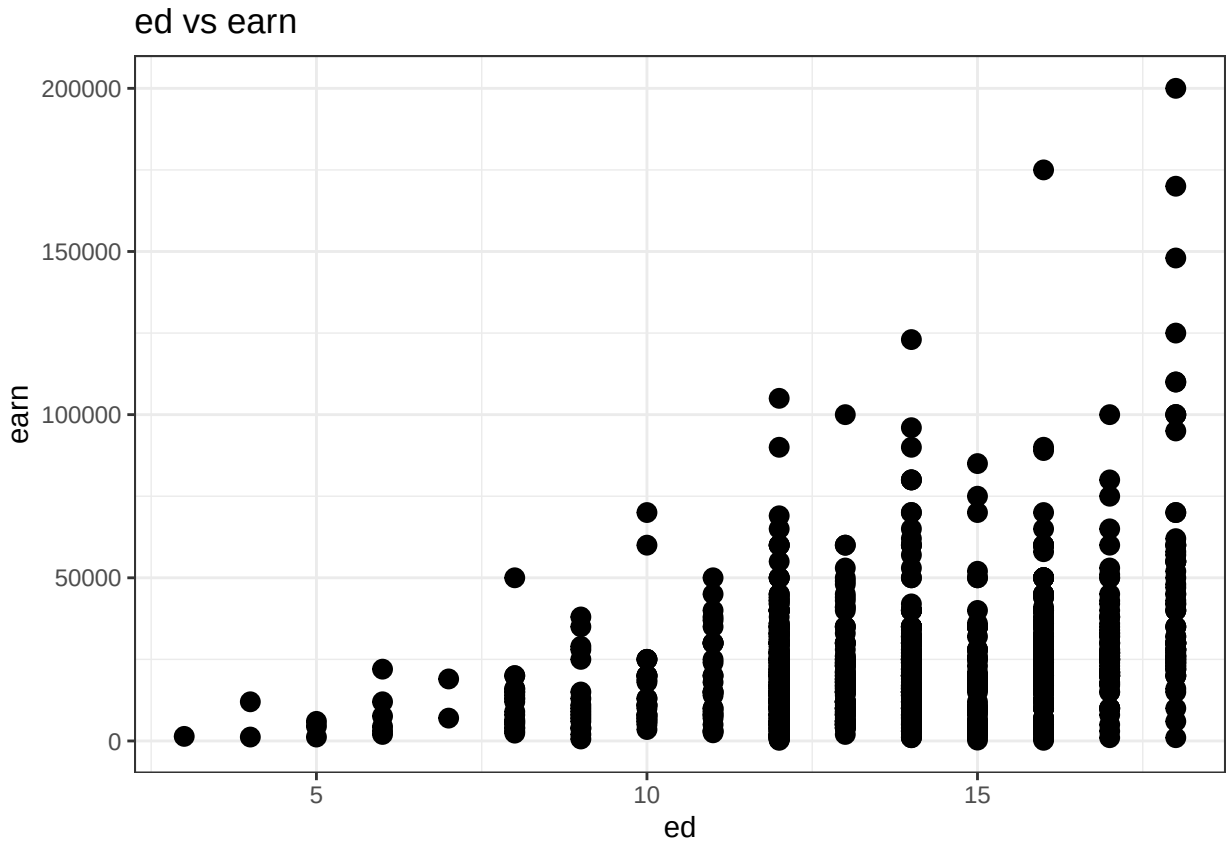
1.5 age vs. earn

```
ggplot(data = heights_df, aes(x = age, y = earn)) + geom_point(size = 3) +
    theme_bw() + labs(title = "age vs earn")
```



1.6 ed vs. earn

```
ggplot(data = heights_df, aes(x = ed, y = earn)) + geom_point(size = 3) +
    theme_bw() + labs(title = "ed vs earn")
```



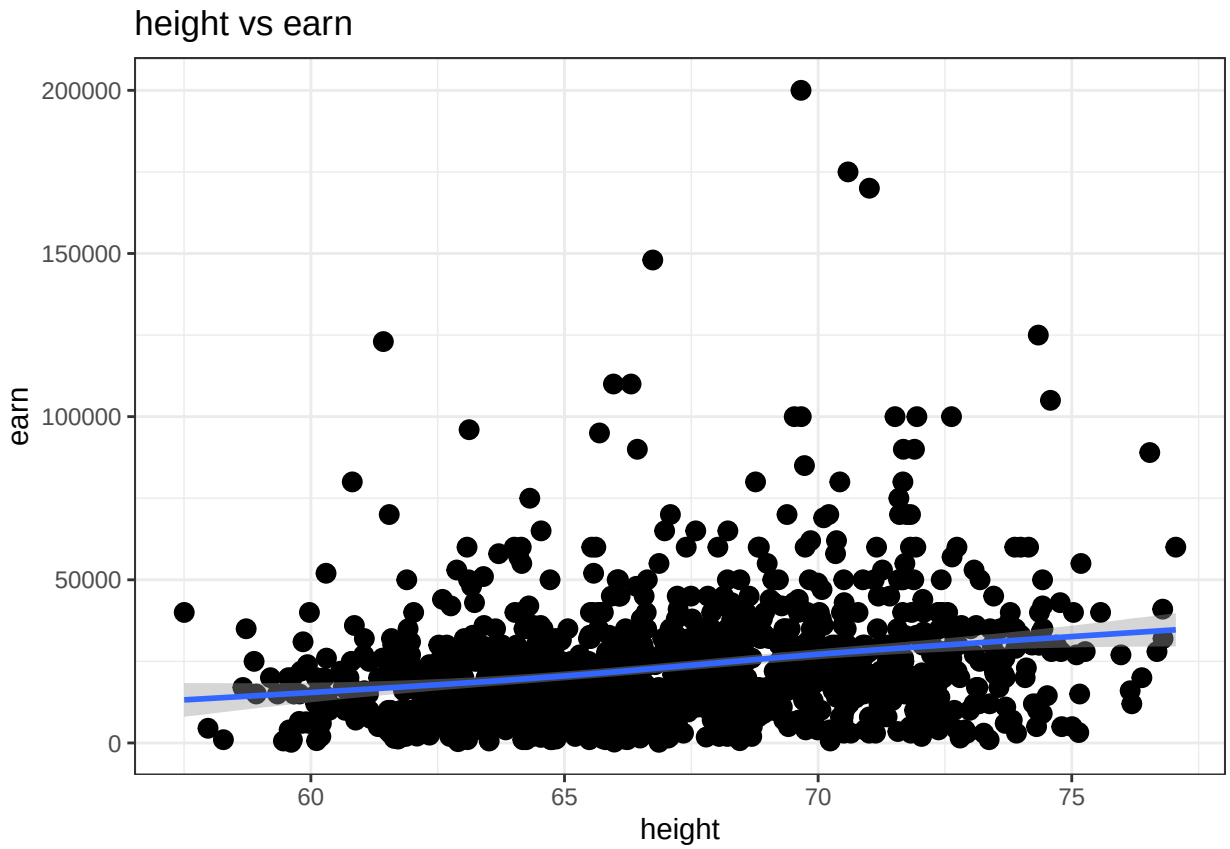
1.7 Re-create the three scatterplots and add a regression trend line using

1.7.1 the geom_smooth() function

1.7.2 height vs. earn

```
ggplot(data = heights_df, aes(x = height, y = earn)) + geom_point(size = 3) +
    geom_smooth() + theme_bw() + labs(title = "height vs earn")
```

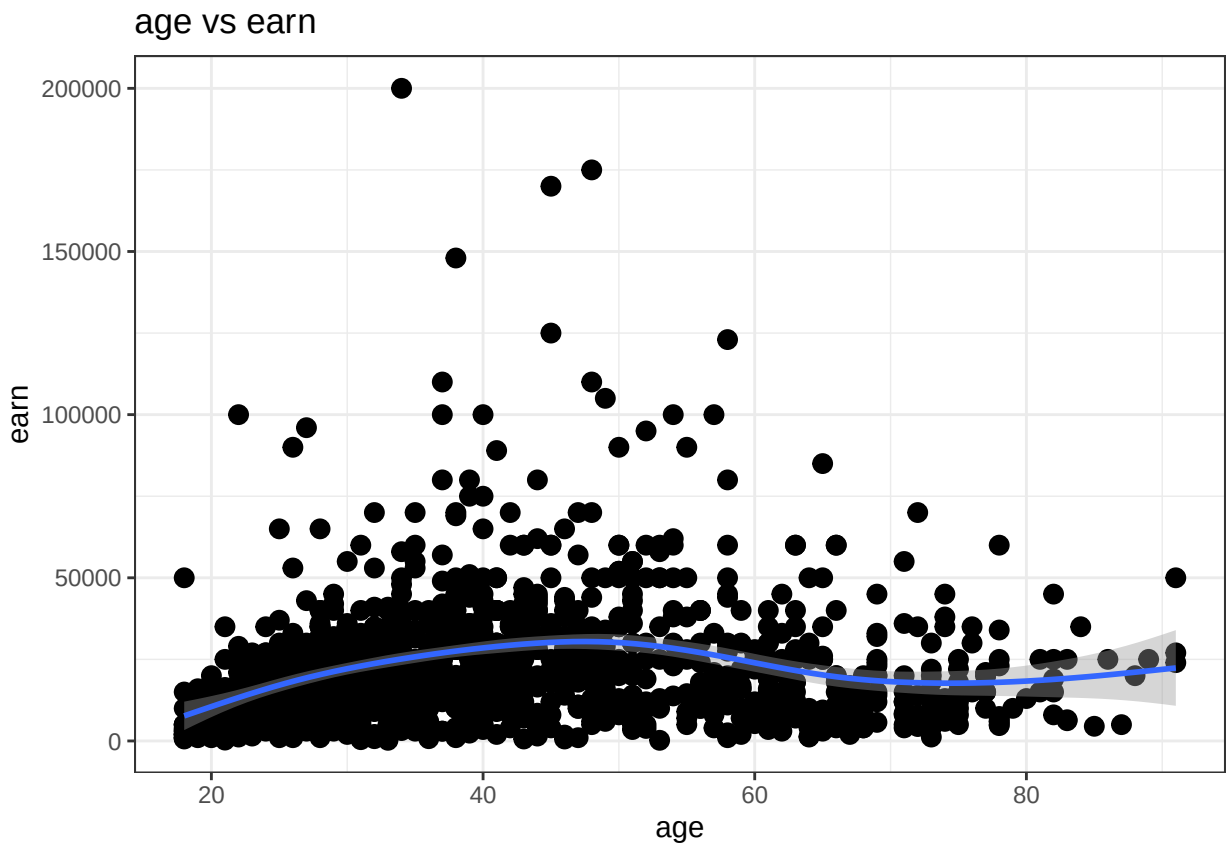
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



1.8 age vs. earn

```
ggplot(data = heights_df, aes(x = age, y = earn)) + geom_point(size = 3) +
    geom_smooth() + theme_bw() + labs(title = "age vs earn")
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

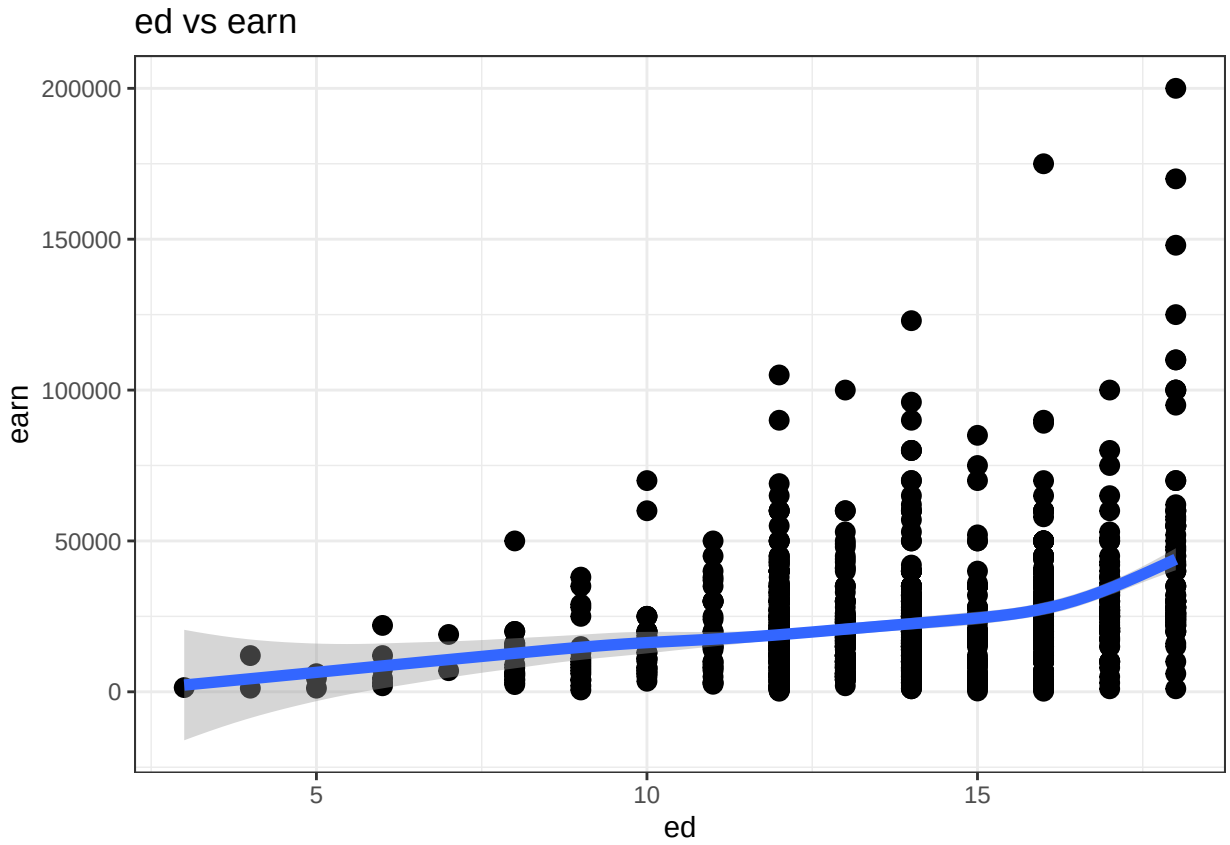


1.9 ed vs. earn

```
ggplot(data = heights_df, aes(x = ed, y = earn)) + geom_point(size = 3) +
    geom_smooth(size = 2) + theme_bw() + labs(title = "ed vs earn")
```

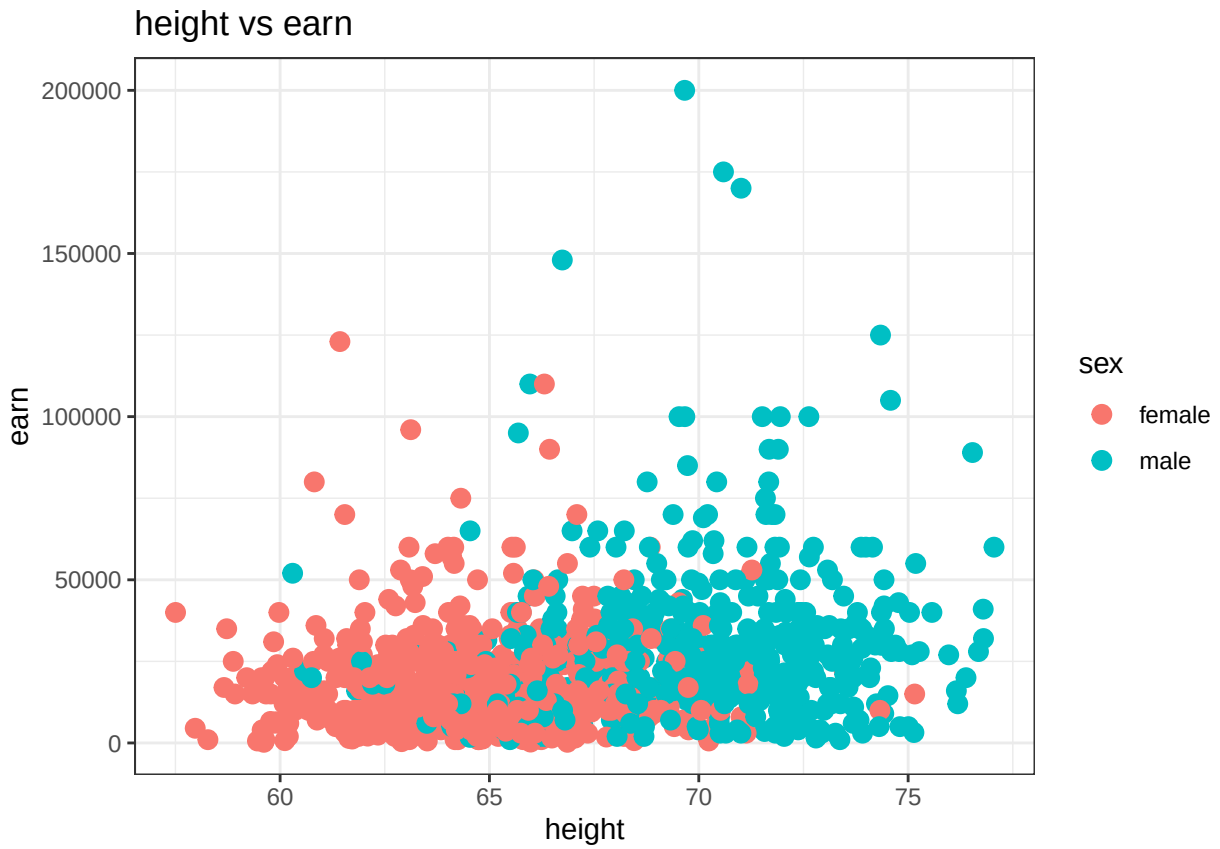
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
```

```
## ‘geom_smooth()‘ using method = ’gam’ and formula = ’y ~ s(x, bs = "cs")’
```



1.10 Create a scatterplot of `height`` vs.earn. Usesexas thecol‘ (color) attribute

```
ggplot(data = heights_df, aes(x = height, y = earn, col = sex)) +
    geom_point(size = 3) + theme_bw() + labs(title = "height vs earn")
```



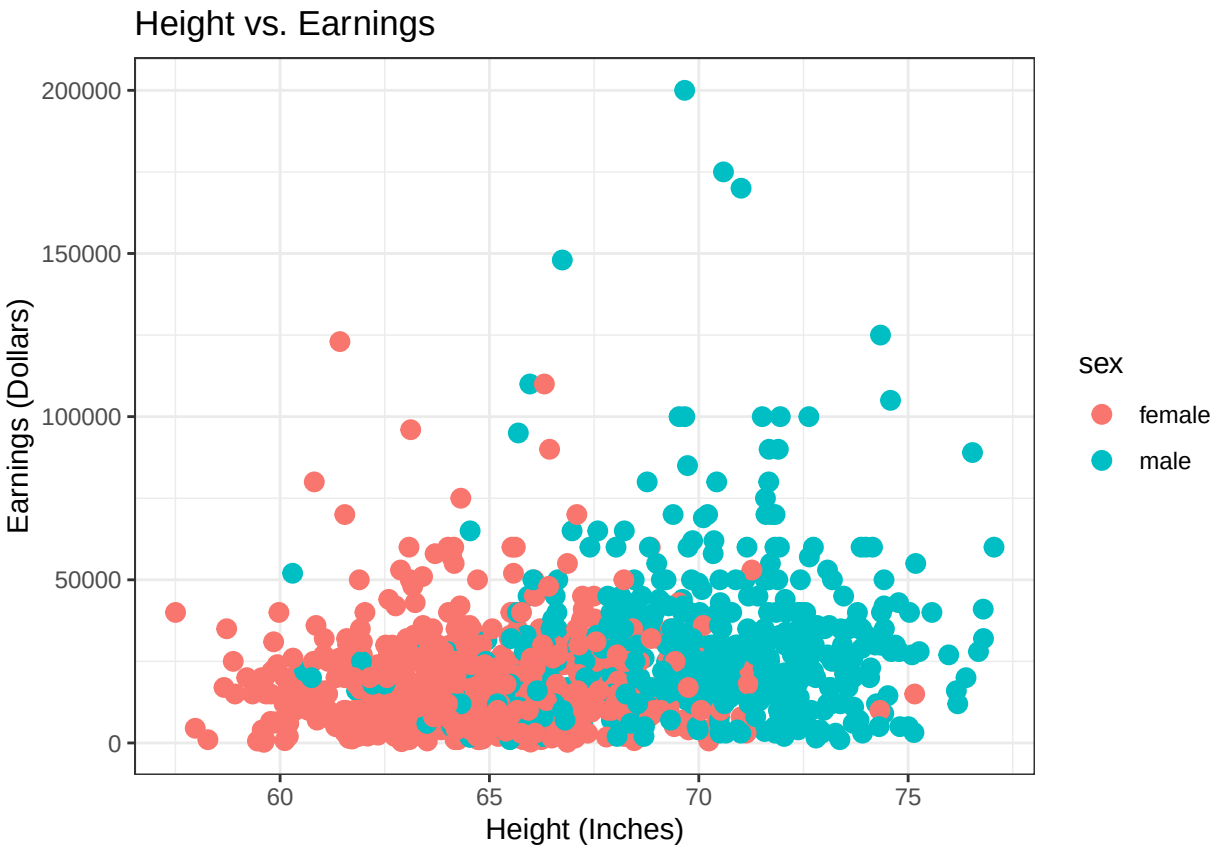
1.11 Using `ggtitle()`, `xlab()`, and `ylab()` to add a title, x label, and y label to the previous plot

1.11.1 Title: Height vs. Earnings

1.11.2 X label: Height (Inches)

1.11.3 Y Label: Earnings (Dollars)

```
ggplot(data = heights_df, aes(x = height, y = earn, col = sex)) +
    geom_point(size = 3) + ggtitle("Height vs. Earnings") + xlab("Height (Inches)") +
    ylab("Earnings (Dollars)") + theme_bw()
```

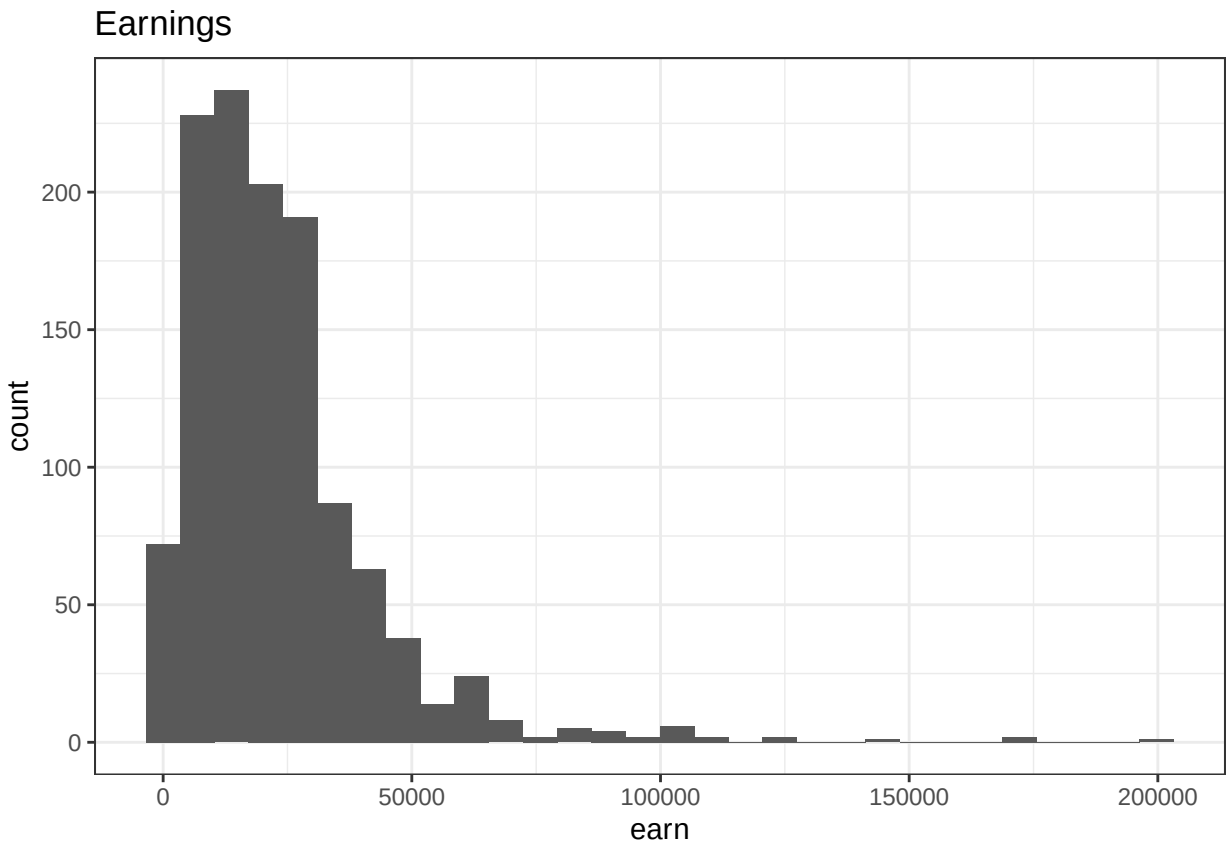


1.12 https://ggplot2.tidyverse.org/reference/geom_histogram.html

1.12.1 Create a histogram of the earn variable using geom_histogram()

```
ggplot(data = heights_df, aes(x = earn)) + geom_histogram() +
    theme_bw() + labs(title = "Earnings")
```

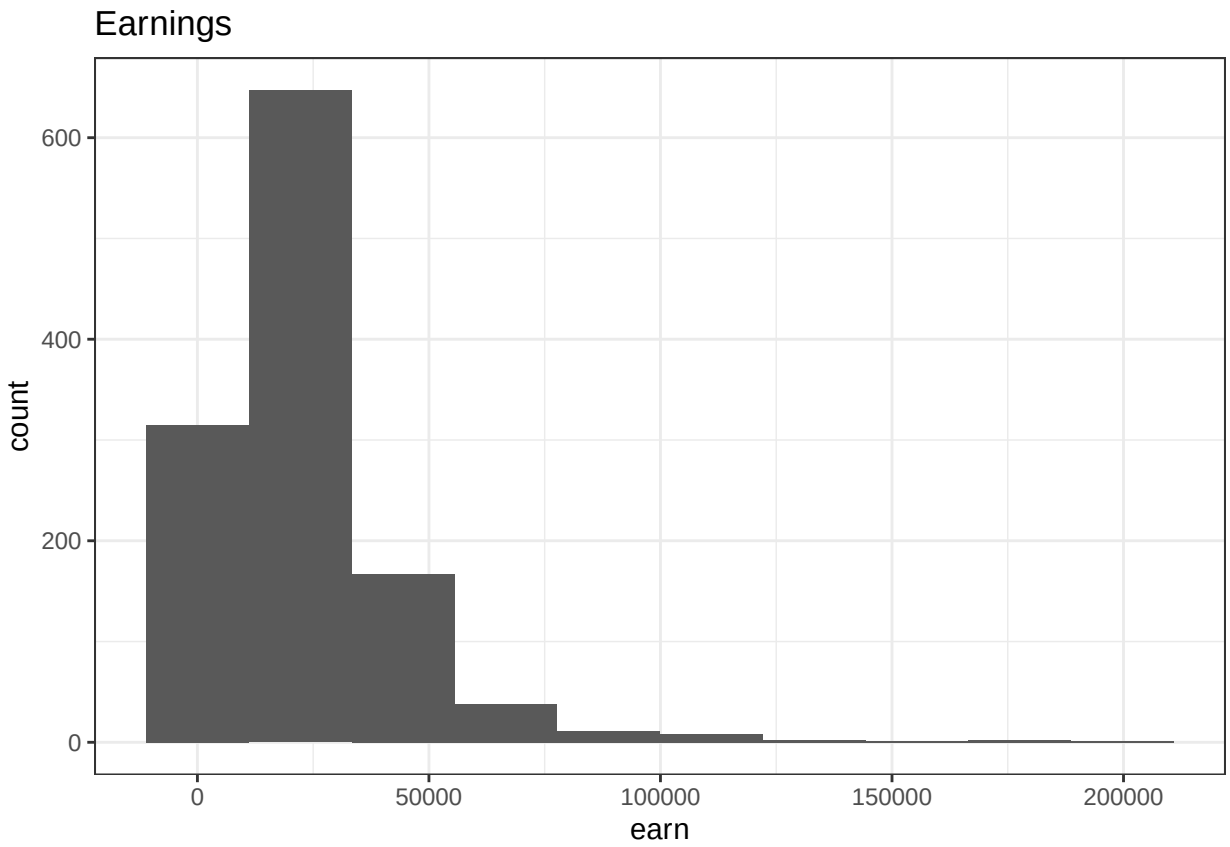
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



1.13 Create a histogram of the earn variable using geom_histogram()

1.13.1 Use 10 bins

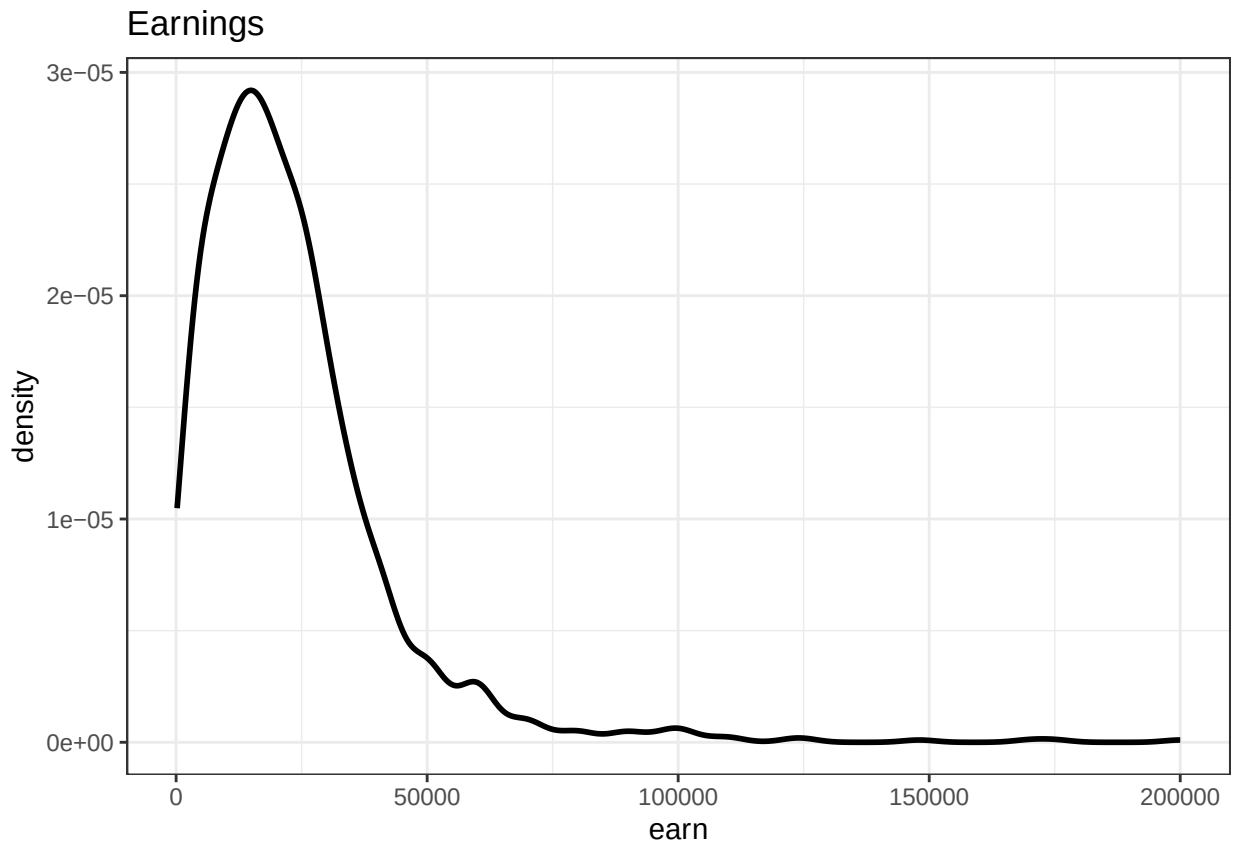
```
ggplot(data = heights_df, aes(x = earn)) + geom_histogram(bins = 10) +
    theme_bw() + labs(title = "Earnings")
```



1.14 https://ggplot2.tidyverse.org/reference/geom_density.html

1.14.1 Create a kernel density plot of `earn` using `geom_density()`

```
ggplot(data = heights_df, aes(x = earn)) + geom_density(size = 1) +
    theme_bw() + labs(title = "Earnings")
```



2 Session Info

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 22621)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] formatR_1.12  ggplot2_3.4.0
##
```

```
## loaded via a namespace (and not attached):
##  [1] highr_0.9        pillar_1.8.1     compiler_4.2.2   tools_4.2.2
##  [5] digest_0.6.30    lattice_0.20-45  nlme_3.1-160     evaluate_0.18
##  [9] lifecycle_1.0.3  tibble_3.1.8     gtable_0.3.1     mgcv_1.8-41
## [13] pkgconfig_2.0.3  rlang_1.0.6      Matrix_1.5-1     cli_3.4.1
## [17] DBI_1.1.3        rstudioapi_0.14  yaml_2.3.6       xfun_0.34
## [21] fastmap_1.1.0    withr_2.5.0      stringr_1.4.1    dplyr_1.0.10
## [25] knitr_1.41       generics_0.1.3   vctrs_0.5.0      grid_4.2.2
## [29] tidyselect_1.2.0 glue_1.6.2       R6_2.5.1         fansi_1.0.3
## [33] rmarkdown_2.18   farver_2.1.1     magrittr_2.0.3   splines_4.2.2
## [37] scales_1.2.1     htmltools_0.5.3  assertthat_0.2.1 colorspace_2.0-3
## [41] labeling_0.4.2   utf8_1.2.2       stringi_1.7.8    munsell_0.5.0
```