FinalButler

LaSheeda Butler

12/14/2021

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.5 ✓ purrr 0.3.4  
## ✓ tibble 3.1.2 ✓ dplyr 1.0.7  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.1

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

library(stringr)   
library(dplyr)  
library(ggplot2)   
library(tidyr)   
library(reshape2)

##   
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':  
##   
## smiths

library(readr)   
library(forcats)  
#1)Using appropriate R code, read in the emailed excel spread sheet Collegedata.   
#Use read\_csv(“Collegedata.csv”)  
read\_csv("Collegedata.csv")->college

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## UNITID = col\_double(),  
## OPEID = col\_double(),  
## MN\_EARN\_WNE\_P6 = col\_character(),  
## INSTNM = col\_character(),  
## SAT\_AVG = col\_double(),  
## ADM\_RATE = col\_double(),  
## UGDS = col\_double(),  
## COSTT4\_A = col\_double(),  
## AVGFACSAL = col\_double(),  
## GRAD\_DEBT\_MDN = col\_character(),  
## AGE\_ENTRY = col\_character(),  
## ICLEVEL = col\_double()  
## )

college

## # A tibble: 7,175 x 12  
## UNITID OPEID MN\_EARN\_WNE\_P6 INSTNM SAT\_AVG ADM\_RATE UGDS COSTT4\_A AVGFACSAL  
## <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 100654 1.00e5 27800 Alaba… 849 0.874 4616 22667 7028  
## 2 100663 1.05e5 37600 Unive… 1125 0.581 12047 22684 10517  
## 3 100690 2.50e6 39400 Amrid… NA NA 293 13380 3857  
## 4 100706 1.06e5 41300 Unive… 1257 0.763 6346 22059 9463  
## 5 100724 1.00e5 23500 Alaba… 825 0.459 4704 19242 7952  
## 6 100751 1.05e5 38500 The U… 1202 0.526 31663 28422 9802  
## 7 100760 1.01e5 26000 Centr… NA NA 1492 13868 5960  
## 8 100812 1.01e5 37400 Athen… NA NA 2888 NA 8367  
## 9 100830 8.31e5 33200 Aubur… 1009 0.766 4171 19255 7251  
## 10 100858 1.01e5 41700 Aubur… 1217 0.805 22095 29794 9945  
## # … with 7,165 more rows, and 3 more variables: GRAD\_DEBT\_MDN <chr>,  
## # AGE\_ENTRY <chr>, ICLEVEL <dbl>

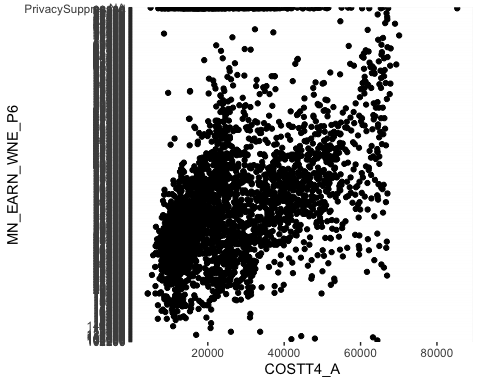
#2)Given the level of the institution, does there appear to be an association   
#between the average cost of attendance(x variable) and the mean earnings of   
#students six years after graduation(y variable)? Make an appropriate plot to   
#justify your response. You will be evaluated on the appropriateness of the plot   
#and the aesthetics of the plot. (Hint: Generate two plots to make your decision,   
#first a standard scatter plot involving the two continuous variables mentioned   
#and then a facet plot over the appropriate categorical variable)  
ggplot(data=college) +  
 geom\_point(mapping = aes(x = `COSTT4\_A`, y=`MN\_EARN\_WNE\_P6`)) +  
 geom\_smooth(mapping = aes(x = `COSTT4\_A`, y=`MN\_EARN\_WNE\_P6`))

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

## Warning: Removed 3520 rows containing non-finite values (stat\_smooth).

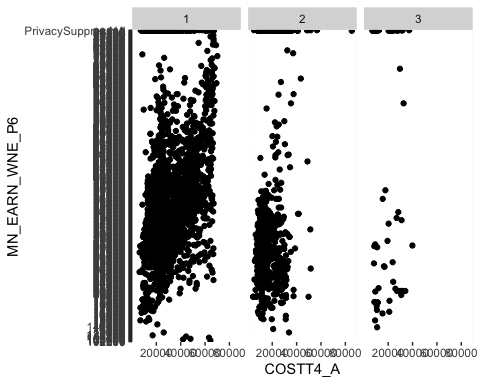
## Warning: Computation failed in `stat\_smooth()`:  
## x has insufficient unique values to support 10 knots: reduce k.

## Warning: Removed 3520 rows containing missing values (geom\_point).



#based on current model, it shows positive association  
  
ggplot(data=college, mapping=(aes(x=`COSTT4\_A`,y=`MN\_EARN\_WNE\_P6`))) +  
 geom\_point() +  
 facet\_wrap(~`ICLEVEL`)

## Warning: Removed 3520 rows containing missing values (geom\_point).



#Looking at the facet plot, it appears that less than 3 years has no assaociation with income.  
  
  
#3)Use r code to produce a histogram of the average age of entry. Comment on  
#the distribution of this variable.  
mutate(college, `AGE\_ENTRY` =parse\_number(`AGE\_ENTRY`))->age

## Warning: 137 parsing failures.  
## row col expected actual  
## 192 -- a number PrivacySuppressed  
## 193 -- a number PrivacySuppressed  
## 196 -- a number PrivacySuppressed  
## 216 -- a number PrivacySuppressed  
## 243 -- a number PrivacySuppressed  
## ... ... ........ .................  
## See problems(...) for more details.

age

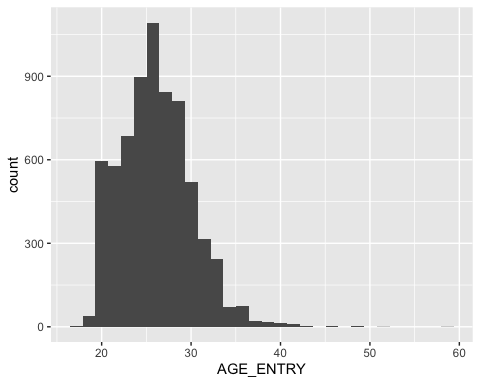
## # A tibble: 7,175 x 12  
## UNITID OPEID MN\_EARN\_WNE\_P6 INSTNM SAT\_AVG ADM\_RATE UGDS COSTT4\_A AVGFACSAL  
## <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 100654 1.00e5 27800 Alaba… 849 0.874 4616 22667 7028  
## 2 100663 1.05e5 37600 Unive… 1125 0.581 12047 22684 10517  
## 3 100690 2.50e6 39400 Amrid… NA NA 293 13380 3857  
## 4 100706 1.06e5 41300 Unive… 1257 0.763 6346 22059 9463  
## 5 100724 1.00e5 23500 Alaba… 825 0.459 4704 19242 7952  
## 6 100751 1.05e5 38500 The U… 1202 0.526 31663 28422 9802  
## 7 100760 1.01e5 26000 Centr… NA NA 1492 13868 5960  
## 8 100812 1.01e5 37400 Athen… NA NA 2888 NA 8367  
## 9 100830 8.31e5 33200 Aubur… 1009 0.766 4171 19255 7251  
## 10 100858 1.01e5 41700 Aubur… 1217 0.805 22095 29794 9945  
## # … with 7,165 more rows, and 3 more variables: GRAD\_DEBT\_MDN <chr>,  
## # AGE\_ENTRY <dbl>, ICLEVEL <dbl>

ggplot(data = age) +  
 geom\_histogram(mapping = aes(x=`AGE\_ENTRY`, stat="count"))

## Warning: Ignoring unknown aesthetics: stat

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

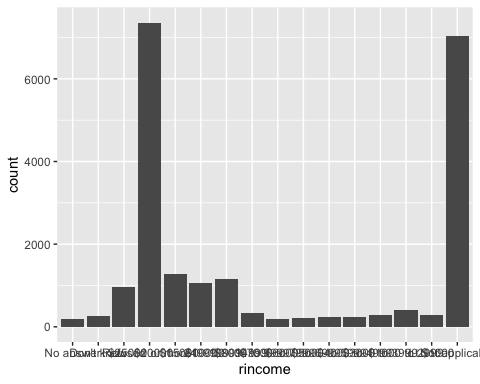
## Warning: Removed 336 rows containing non-finite values (stat\_bin).



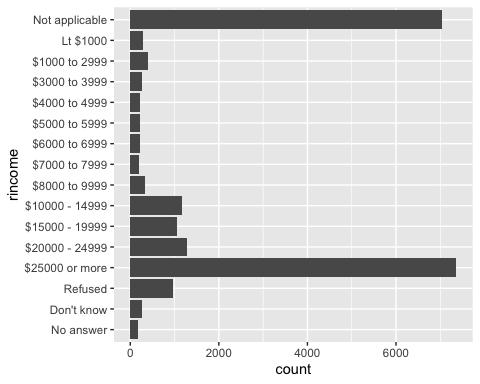
#4)Use r code that will produce output that shows the 10 institutions that have   
#the highest average age of entry?  
#age%>%  
 #sort(desc(AGE\_ENTRY))%>%  
 #print(n=10)  
   
  
   
#5)There are many universities with "American University" in the name. E.g.  
#"American University of Puerto Rico" and "National American  
# University-Ellsworth AFB Extension". Use R code to create a data frame, called   
#`americandf`, that contains just the data from universities with  
#"American University" in the name.  
  
  
  
  
#6)Provide r code that will produce the number of colleges from the College   
#Score data frame that have average SAT scores that are above 1000. ( Do not   
#produce the data frame. Your code should only yield the number)  
college%>%  
 filter(`SAT\_AVG`>=1000)%>%  
 count()

## # A tibble: 1 x 1  
## n  
## <int>  
## 1 852

#7)Provide r code that will show a data frame that lists the 10 highest Average   
#SAT scores in decreasing order. A partial data frame is given below.  
#college%>%  
 #select(UNITID, OPEID, MN\_EARN\_WNE\_P6, INSTNM, SAT\_AVG, ADM\_RATE, UGDS, COSTT4\_A)  
 # filter(`SAT\_AVG`>=1000)->highsat  
#highsat  
  
#highsat%>%  
 #sort(desc(`SAT\_AVG`))  
  
#8) Using the gss\_cat data frame, write r code that will produce the bar graph   
#below. And explain in one or two sentences why the bar graph is difficult to interpret.  
ggplot(data=gss\_cat) +  
 geom\_bar(mapping = aes(rincome))



#It's difficult to interpret the data because the numbers are extremely close together.  
  
  
  
#9 Now write r code from the same data set that produce the transformed bar graph   
#and comment on why it is an improvement  
ggplot(data=gss\_cat) +  
 geom\_bar(mapping = aes(y=rincome))



#9)Use r code to produce the tips data frame from the reshape2 package. Name   
#three categorical variables in the data frame.  
tips

## total\_bill tip sex smoker day time size  
## 1 16.99 1.01 Female No Sun Dinner 2  
## 2 10.34 1.66 Male No Sun Dinner 3  
## 3 21.01 3.50 Male No Sun Dinner 3  
## 4 23.68 3.31 Male No Sun Dinner 2  
## 5 24.59 3.61 Female No Sun Dinner 4  
## 6 25.29 4.71 Male No Sun Dinner 4  
## 7 8.77 2.00 Male No Sun Dinner 2  
## 8 26.88 3.12 Male No Sun Dinner 4  
## 9 15.04 1.96 Male No Sun Dinner 2  
## 10 14.78 3.23 Male No Sun Dinner 2  
## 11 10.27 1.71 Male No Sun Dinner 2  
## 12 35.26 5.00 Female No Sun Dinner 4  
## 13 15.42 1.57 Male No Sun Dinner 2  
## 14 18.43 3.00 Male No Sun Dinner 4  
## 15 14.83 3.02 Female No Sun Dinner 2  
## 16 21.58 3.92 Male No Sun Dinner 2  
## 17 10.33 1.67 Female No Sun Dinner 3  
## 18 16.29 3.71 Male No Sun Dinner 3  
## 19 16.97 3.50 Female No Sun Dinner 3  
## 20 20.65 3.35 Male No Sat Dinner 3  
## 21 17.92 4.08 Male No Sat Dinner 2  
## 22 20.29 2.75 Female No Sat Dinner 2  
## 23 15.77 2.23 Female No Sat Dinner 2  
## 24 39.42 7.58 Male No Sat Dinner 4  
## 25 19.82 3.18 Male No Sat Dinner 2  
## 26 17.81 2.34 Male No Sat Dinner 4  
## 27 13.37 2.00 Male No Sat Dinner 2  
## 28 12.69 2.00 Male No Sat Dinner 2  
## 29 21.70 4.30 Male No Sat Dinner 2  
## 30 19.65 3.00 Female No Sat Dinner 2  
## 31 9.55 1.45 Male No Sat Dinner 2  
## 32 18.35 2.50 Male No Sat Dinner 4  
## 33 15.06 3.00 Female No Sat Dinner 2  
## 34 20.69 2.45 Female No Sat Dinner 4  
## 35 17.78 3.27 Male No Sat Dinner 2  
## 36 24.06 3.60 Male No Sat Dinner 3  
## 37 16.31 2.00 Male No Sat Dinner 3  
## 38 16.93 3.07 Female No Sat Dinner 3  
## 39 18.69 2.31 Male No Sat Dinner 3  
## 40 31.27 5.00 Male No Sat Dinner 3  
## 41 16.04 2.24 Male No Sat Dinner 3  
## 42 17.46 2.54 Male No Sun Dinner 2  
## 43 13.94 3.06 Male No Sun Dinner 2  
## 44 9.68 1.32 Male No Sun Dinner 2  
## 45 30.40 5.60 Male No Sun Dinner 4  
## 46 18.29 3.00 Male No Sun Dinner 2  
## 47 22.23 5.00 Male No Sun Dinner 2  
## 48 32.40 6.00 Male No Sun Dinner 4  
## 49 28.55 2.05 Male No Sun Dinner 3  
## 50 18.04 3.00 Male No Sun Dinner 2  
## 51 12.54 2.50 Male No Sun Dinner 2  
## 52 10.29 2.60 Female No Sun Dinner 2  
## 53 34.81 5.20 Female No Sun Dinner 4  
## 54 9.94 1.56 Male No Sun Dinner 2  
## 55 25.56 4.34 Male No Sun Dinner 4  
## 56 19.49 3.51 Male No Sun Dinner 2  
## 57 38.01 3.00 Male Yes Sat Dinner 4  
## 58 26.41 1.50 Female No Sat Dinner 2  
## 59 11.24 1.76 Male Yes Sat Dinner 2  
## 60 48.27 6.73 Male No Sat Dinner 4  
## 61 20.29 3.21 Male Yes Sat Dinner 2  
## 62 13.81 2.00 Male Yes Sat Dinner 2  
## 63 11.02 1.98 Male Yes Sat Dinner 2  
## 64 18.29 3.76 Male Yes Sat Dinner 4  
## 65 17.59 2.64 Male No Sat Dinner 3  
## 66 20.08 3.15 Male No Sat Dinner 3  
## 67 16.45 2.47 Female No Sat Dinner 2  
## 68 3.07 1.00 Female Yes Sat Dinner 1  
## 69 20.23 2.01 Male No Sat Dinner 2  
## 70 15.01 2.09 Male Yes Sat Dinner 2  
## 71 12.02 1.97 Male No Sat Dinner 2  
## 72 17.07 3.00 Female No Sat Dinner 3  
## 73 26.86 3.14 Female Yes Sat Dinner 2  
## 74 25.28 5.00 Female Yes Sat Dinner 2  
## 75 14.73 2.20 Female No Sat Dinner 2  
## 76 10.51 1.25 Male No Sat Dinner 2  
## 77 17.92 3.08 Male Yes Sat Dinner 2  
## 78 27.20 4.00 Male No Thur Lunch 4  
## 79 22.76 3.00 Male No Thur Lunch 2  
## 80 17.29 2.71 Male No Thur Lunch 2  
## 81 19.44 3.00 Male Yes Thur Lunch 2  
## 82 16.66 3.40 Male No Thur Lunch 2  
## 83 10.07 1.83 Female No Thur Lunch 1  
## 84 32.68 5.00 Male Yes Thur Lunch 2  
## 85 15.98 2.03 Male No Thur Lunch 2  
## 86 34.83 5.17 Female No Thur Lunch 4  
## 87 13.03 2.00 Male No Thur Lunch 2  
## 88 18.28 4.00 Male No Thur Lunch 2  
## 89 24.71 5.85 Male No Thur Lunch 2  
## 90 21.16 3.00 Male No Thur Lunch 2  
## 91 28.97 3.00 Male Yes Fri Dinner 2  
## 92 22.49 3.50 Male No Fri Dinner 2  
## 93 5.75 1.00 Female Yes Fri Dinner 2  
## 94 16.32 4.30 Female Yes Fri Dinner 2  
## 95 22.75 3.25 Female No Fri Dinner 2  
## 96 40.17 4.73 Male Yes Fri Dinner 4  
## 97 27.28 4.00 Male Yes Fri Dinner 2  
## 98 12.03 1.50 Male Yes Fri Dinner 2  
## 99 21.01 3.00 Male Yes Fri Dinner 2  
## 100 12.46 1.50 Male No Fri Dinner 2  
## 101 11.35 2.50 Female Yes Fri Dinner 2  
## 102 15.38 3.00 Female Yes Fri Dinner 2  
## 103 44.30 2.50 Female Yes Sat Dinner 3  
## 104 22.42 3.48 Female Yes Sat Dinner 2  
## 105 20.92 4.08 Female No Sat Dinner 2  
## 106 15.36 1.64 Male Yes Sat Dinner 2  
## 107 20.49 4.06 Male Yes Sat Dinner 2  
## 108 25.21 4.29 Male Yes Sat Dinner 2  
## 109 18.24 3.76 Male No Sat Dinner 2  
## 110 14.31 4.00 Female Yes Sat Dinner 2  
## 111 14.00 3.00 Male No Sat Dinner 2  
## 112 7.25 1.00 Female No Sat Dinner 1  
## 113 38.07 4.00 Male No Sun Dinner 3  
## 114 23.95 2.55 Male No Sun Dinner 2  
## 115 25.71 4.00 Female No Sun Dinner 3  
## 116 17.31 3.50 Female No Sun Dinner 2  
## 117 29.93 5.07 Male No Sun Dinner 4  
## 118 10.65 1.50 Female No Thur Lunch 2  
## 119 12.43 1.80 Female No Thur Lunch 2  
## 120 24.08 2.92 Female No Thur Lunch 4  
## 121 11.69 2.31 Male No Thur Lunch 2  
## 122 13.42 1.68 Female No Thur Lunch 2  
## 123 14.26 2.50 Male No Thur Lunch 2  
## 124 15.95 2.00 Male No Thur Lunch 2  
## 125 12.48 2.52 Female No Thur Lunch 2  
## 126 29.80 4.20 Female No Thur Lunch 6  
## 127 8.52 1.48 Male No Thur Lunch 2  
## 128 14.52 2.00 Female No Thur Lunch 2  
## 129 11.38 2.00 Female No Thur Lunch 2  
## 130 22.82 2.18 Male No Thur Lunch 3  
## 131 19.08 1.50 Male No Thur Lunch 2  
## 132 20.27 2.83 Female No Thur Lunch 2  
## 133 11.17 1.50 Female No Thur Lunch 2  
## 134 12.26 2.00 Female No Thur Lunch 2  
## 135 18.26 3.25 Female No Thur Lunch 2  
## 136 8.51 1.25 Female No Thur Lunch 2  
## 137 10.33 2.00 Female No Thur Lunch 2  
## 138 14.15 2.00 Female No Thur Lunch 2  
## 139 16.00 2.00 Male Yes Thur Lunch 2  
## 140 13.16 2.75 Female No Thur Lunch 2  
## 141 17.47 3.50 Female No Thur Lunch 2  
## 142 34.30 6.70 Male No Thur Lunch 6  
## 143 41.19 5.00 Male No Thur Lunch 5  
## 144 27.05 5.00 Female No Thur Lunch 6  
## 145 16.43 2.30 Female No Thur Lunch 2  
## 146 8.35 1.50 Female No Thur Lunch 2  
## 147 18.64 1.36 Female No Thur Lunch 3  
## 148 11.87 1.63 Female No Thur Lunch 2  
## 149 9.78 1.73 Male No Thur Lunch 2  
## 150 7.51 2.00 Male No Thur Lunch 2  
## 151 14.07 2.50 Male No Sun Dinner 2  
## 152 13.13 2.00 Male No Sun Dinner 2  
## 153 17.26 2.74 Male No Sun Dinner 3  
## 154 24.55 2.00 Male No Sun Dinner 4  
## 155 19.77 2.00 Male No Sun Dinner 4  
## 156 29.85 5.14 Female No Sun Dinner 5  
## 157 48.17 5.00 Male No Sun Dinner 6  
## 158 25.00 3.75 Female No Sun Dinner 4  
## 159 13.39 2.61 Female No Sun Dinner 2  
## 160 16.49 2.00 Male No Sun Dinner 4  
## 161 21.50 3.50 Male No Sun Dinner 4  
## 162 12.66 2.50 Male No Sun Dinner 2  
## 163 16.21 2.00 Female No Sun Dinner 3  
## 164 13.81 2.00 Male No Sun Dinner 2  
## 165 17.51 3.00 Female Yes Sun Dinner 2  
## 166 24.52 3.48 Male No Sun Dinner 3  
## 167 20.76 2.24 Male No Sun Dinner 2  
## 168 31.71 4.50 Male No Sun Dinner 4  
## 169 10.59 1.61 Female Yes Sat Dinner 2  
## 170 10.63 2.00 Female Yes Sat Dinner 2  
## 171 50.81 10.00 Male Yes Sat Dinner 3  
## 172 15.81 3.16 Male Yes Sat Dinner 2  
## 173 7.25 5.15 Male Yes Sun Dinner 2  
## 174 31.85 3.18 Male Yes Sun Dinner 2  
## 175 16.82 4.00 Male Yes Sun Dinner 2  
## 176 32.90 3.11 Male Yes Sun Dinner 2  
## 177 17.89 2.00 Male Yes Sun Dinner 2  
## 178 14.48 2.00 Male Yes Sun Dinner 2  
## 179 9.60 4.00 Female Yes Sun Dinner 2  
## 180 34.63 3.55 Male Yes Sun Dinner 2  
## 181 34.65 3.68 Male Yes Sun Dinner 4  
## 182 23.33 5.65 Male Yes Sun Dinner 2  
## 183 45.35 3.50 Male Yes Sun Dinner 3  
## 184 23.17 6.50 Male Yes Sun Dinner 4  
## 185 40.55 3.00 Male Yes Sun Dinner 2  
## 186 20.69 5.00 Male No Sun Dinner 5  
## 187 20.90 3.50 Female Yes Sun Dinner 3  
## 188 30.46 2.00 Male Yes Sun Dinner 5  
## 189 18.15 3.50 Female Yes Sun Dinner 3  
## 190 23.10 4.00 Male Yes Sun Dinner 3  
## 191 15.69 1.50 Male Yes Sun Dinner 2  
## 192 19.81 4.19 Female Yes Thur Lunch 2  
## 193 28.44 2.56 Male Yes Thur Lunch 2  
## 194 15.48 2.02 Male Yes Thur Lunch 2  
## 195 16.58 4.00 Male Yes Thur Lunch 2  
## 196 7.56 1.44 Male No Thur Lunch 2  
## 197 10.34 2.00 Male Yes Thur Lunch 2  
## 198 43.11 5.00 Female Yes Thur Lunch 4  
## 199 13.00 2.00 Female Yes Thur Lunch 2  
## 200 13.51 2.00 Male Yes Thur Lunch 2  
## 201 18.71 4.00 Male Yes Thur Lunch 3  
## 202 12.74 2.01 Female Yes Thur Lunch 2  
## 203 13.00 2.00 Female Yes Thur Lunch 2  
## 204 16.40 2.50 Female Yes Thur Lunch 2  
## 205 20.53 4.00 Male Yes Thur Lunch 4  
## 206 16.47 3.23 Female Yes Thur Lunch 3  
## 207 26.59 3.41 Male Yes Sat Dinner 3  
## 208 38.73 3.00 Male Yes Sat Dinner 4  
## 209 24.27 2.03 Male Yes Sat Dinner 2  
## 210 12.76 2.23 Female Yes Sat Dinner 2  
## 211 30.06 2.00 Male Yes Sat Dinner 3  
## 212 25.89 5.16 Male Yes Sat Dinner 4  
## 213 48.33 9.00 Male No Sat Dinner 4  
## 214 13.27 2.50 Female Yes Sat Dinner 2  
## 215 28.17 6.50 Female Yes Sat Dinner 3  
## 216 12.90 1.10 Female Yes Sat Dinner 2  
## 217 28.15 3.00 Male Yes Sat Dinner 5  
## 218 11.59 1.50 Male Yes Sat Dinner 2  
## 219 7.74 1.44 Male Yes Sat Dinner 2  
## 220 30.14 3.09 Female Yes Sat Dinner 4  
## 221 12.16 2.20 Male Yes Fri Lunch 2  
## 222 13.42 3.48 Female Yes Fri Lunch 2  
## 223 8.58 1.92 Male Yes Fri Lunch 1  
## 224 15.98 3.00 Female No Fri Lunch 3  
## 225 13.42 1.58 Male Yes Fri Lunch 2  
## 226 16.27 2.50 Female Yes Fri Lunch 2  
## 227 10.09 2.00 Female Yes Fri Lunch 2  
## 228 20.45 3.00 Male No Sat Dinner 4  
## 229 13.28 2.72 Male No Sat Dinner 2  
## 230 22.12 2.88 Female Yes Sat Dinner 2  
## 231 24.01 2.00 Male Yes Sat Dinner 4  
## 232 15.69 3.00 Male Yes Sat Dinner 3  
## 233 11.61 3.39 Male No Sat Dinner 2  
## 234 10.77 1.47 Male No Sat Dinner 2  
## 235 15.53 3.00 Male Yes Sat Dinner 2  
## 236 10.07 1.25 Male No Sat Dinner 2  
## 237 12.60 1.00 Male Yes Sat Dinner 2  
## 238 32.83 1.17 Male Yes Sat Dinner 2  
## 239 35.83 4.67 Female No Sat Dinner 3  
## 240 29.03 5.92 Male No Sat Dinner 3  
## 241 27.18 2.00 Female Yes Sat Dinner 2  
## 242 22.67 2.00 Male Yes Sat Dinner 2  
## 243 17.82 1.75 Male No Sat Dinner 2  
## 244 18.78 3.00 Female No Thur Dinner 2

#sex, worker, day  
  
#10)Use r code to indicate how many levels exist for the factor day in the tips   
#data frame and determine the frequency of each level.  
levels(tips$day)

## [1] "Fri" "Sat" "Sun" "Thur"

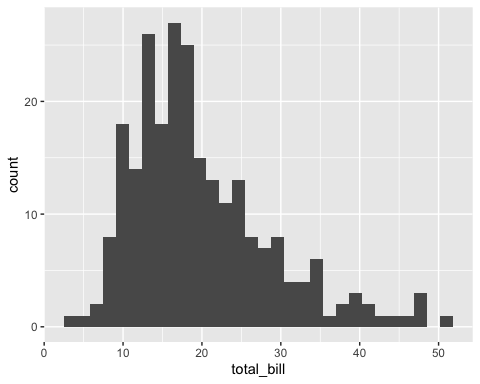
tips%>%  
 count(day)

## day n  
## 1 Fri 19  
## 2 Sat 87  
## 3 Sun 76  
## 4 Thur 62

#11)Produce r code that will produce the following histogram from the tips data frame  
ggplot(data = tips) +  
 geom\_histogram(mapping = aes(x=total\_bill, binwidth = 2))

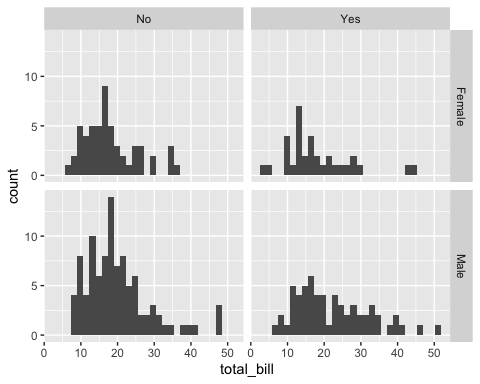
## Warning: Ignoring unknown aesthetics: binwidth

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#12)Write r code that will produce the following histograms from the tips data frame  
  
ggplot(data = tips) +  
 geom\_histogram(mapping= aes(x=total\_bill)) +  
 facet\_grid(sex~smoker)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#13)Using the stringr::words data set along with str\_subset code, produce R code   
#that will show a 9 letter word that has the letter a in the middle  
str\_subset(words, "....a....")

## [1] "character"

str\_subset(words, "^....a....$")

## [1] "character"

#14)Produce a string that will force a match for the regular expression \\””\   
#Use and show the R code command writelines to confirm your answer  
  
  
#15)Describe in words (two or three sentences) what the following regular expression will match ^.\*e$  
str\_subset(words, "^.\*e$")

## [1] "able" "absolute" "achieve" "active" "advertise"   
## [6] "age" "agree" "appropriate" "argue" "arrange"   
## [11] "associate" "assume" "available" "aware" "balance"   
## [16] "base" "be" "because" "become" "before"   
## [21] "believe" "bloke" "blue" "bottle" "cake"   
## [26] "care" "case" "cause" "centre" "chance"   
## [31] "change" "charge" "choice" "choose" "close"   
## [36] "clothe" "coffee" "colleague" "college" "come"   
## [41] "committee" "compare" "complete" "compute" "continue"   
## [46] "converse" "couple" "course" "create" "date"   
## [51] "debate" "decide" "definite" "degree" "describe"   
## [56] "die" "difference" "divide" "double" "drive"   
## [61] "due" "educate" "else" "encourage" "engine"   
## [66] "europe" "evidence" "example" "excuse" "exercise"   
## [71] "expense" "experience" "eye" "face" "figure"   
## [76] "file" "finance" "fine" "fire" "five"   
## [81] "force" "fortune" "france" "free" "future"   
## [86] "game" "give" "goodbye" "hate" "have"   
## [91] "he" "here" "home" "hope" "horse"   
## [96] "house" "imagine" "improve" "include" "income"   
## [101] "increase" "inside" "insure" "introduce" "involve"   
## [106] "issue" "judge" "language" "large" "late"   
## [111] "leave" "lie" "life" "like" "line"   
## [116] "little" "live" "lose" "love" "machine"   
## [121] "make" "manage" "maybe" "measure" "middle"   
## [126] "mile" "minute" "more" "move" "name"   
## [131] "nature" "nice" "nine" "none" "note"   
## [136] "notice" "office" "once" "one" "operate"   
## [141] "oppose" "organize" "otherwise" "page" "pence"   
## [146] "people" "picture" "piece" "place" "please"   
## [151] "police" "positive" "possible" "practise" "prepare"   
## [156] "pressure" "presume" "price" "private" "probable"   
## [161] "produce" "programme" "propose" "provide" "purpose"   
## [166] "quite" "raise" "range" "rate" "realise"   
## [171] "receive" "recognize" "reduce" "require" "resource"   
## [176] "responsible" "rise" "role" "rule" "safe"   
## [181] "sale" "same" "save" "scheme" "science"   
## [186] "score" "secure" "see" "sense" "separate"   
## [191] "serve" "service" "settle" "share" "she"   
## [196] "shoe" "side" "simple" "since" "single"   
## [201] "site" "situate" "size" "smoke" "some"   
## [206] "space" "square" "stage" "state" "strike"   
## [211] "structure" "suppose" "sure" "surprise" "table"   
## [216] "take" "tape" "telephone" "terrible" "the"   
## [221] "there" "therefore" "three" "tie" "time"   
## [226] "trade" "tree" "trouble" "true" "twelve"   
## [231] "type" "unite" "use" "value" "village"   
## [236] "vote" "wage" "waste" "we" "wee"   
## [241] "welcome" "where" "while" "white" "whole"   
## [246] "wide" "wife" "worse" "write"

#This means that a word can start with any letter, but it must end with an "e".  
  
  
#16)Using the methods demonstrated in class regarding Factors and Forcats, use   
#and show R code to create a factor that will enable you to sort the string vector   
#(“eight”, “four”, “ten”, “two”) according to quantity, not alphabetical order.   
r<-c("eight", "four", "ten", "two")   
r

## [1] "eight" "four" "ten" "two"

ones<-c("one", "two", "three", "four", "five", "six", "seven", "eight", "nine" ,"ten")  
ones

## [1] "one" "two" "three" "four" "five" "six" "seven" "eight" "nine"   
## [10] "ten"

r2<-factor(r, levels = ones)  
r2

## [1] eight four ten two   
## Levels: one two three four five six seven eight nine ten

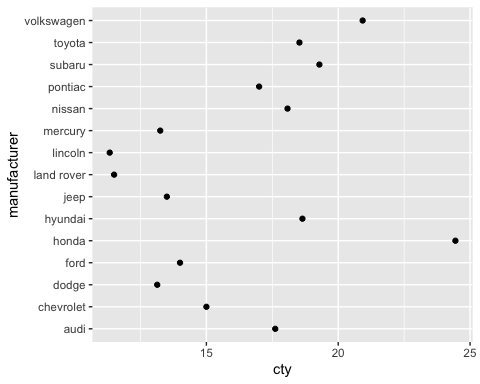
sort(r2)

## [1] two four eight ten   
## Levels: one two three four five six seven eight nine ten

#17)Using the mpg data set, use and show R code that will produce a table that   
#shows the average city mileage (mean for cty) for each manufacturer.  
mpg%>%  
 group\_by(manufacturer)%>%  
 summarize(cty= mean(cty, na.rm=TRUE), n=n())->meancty  
meancty

## # A tibble: 15 x 3  
## manufacturer cty n  
## <chr> <dbl> <int>  
## 1 audi 17.6 18  
## 2 chevrolet 15 19  
## 3 dodge 13.1 37  
## 4 ford 14 25  
## 5 honda 24.4 9  
## 6 hyundai 18.6 14  
## 7 jeep 13.5 8  
## 8 land rover 11.5 4  
## 9 lincoln 11.3 3  
## 10 mercury 13.2 4  
## 11 nissan 18.1 13  
## 12 pontiac 17 5  
## 13 subaru 19.3 14  
## 14 toyota 18.5 34  
## 15 volkswagen 20.9 27

#18)Using the mpg data set and methods and code illustrated in the emailed Factors   
#and Forcats R file: use and show R code to generate a scatter plot that clearly  
#shows how different manufacturers compare with one another with regards to average   
#or mean city mileage (cty)   
ggplot(data = meancty) +  
 geom\_point(mapping = aes(x=cty,  
 y = manufacturer))



#19) To the mpg data table, apply a stringr function to print all observations   
#of the manufacturer variable in upper case letters. Then use R coding to produce   
#rows 15 to 25. The first five rows of the table are shown below. Note that the   
#row numbers 1 – 5 correspond to rows 15 – 20.  
mpg%>%  
 select(manufacturer, model, year)%>%  
 print(n=25)->grr

## # A tibble: 234 x 3  
## manufacturer model year  
## <chr> <chr> <int>  
## 1 audi a4 1999  
## 2 audi a4 1999  
## 3 audi a4 2008  
## 4 audi a4 2008  
## 5 audi a4 1999  
## 6 audi a4 1999  
## 7 audi a4 2008  
## 8 audi a4 quattro 1999  
## 9 audi a4 quattro 1999  
## 10 audi a4 quattro 2008  
## 11 audi a4 quattro 2008  
## 12 audi a4 quattro 1999  
## 13 audi a4 quattro 1999  
## 14 audi a4 quattro 2008  
## 15 audi a4 quattro 2008  
## 16 audi a6 quattro 1999  
## 17 audi a6 quattro 2008  
## 18 audi a6 quattro 2008  
## 19 chevrolet c1500 suburban 2wd 2008  
## 20 chevrolet c1500 suburban 2wd 2008  
## 21 chevrolet c1500 suburban 2wd 2008  
## 22 chevrolet c1500 suburban 2wd 1999  
## 23 chevrolet c1500 suburban 2wd 2008  
## 24 chevrolet corvette 1999  
## 25 chevrolet corvette 1999  
## # … with 209 more rows

grr

## # A tibble: 234 x 3  
## manufacturer model year  
## <chr> <chr> <int>  
## 1 audi a4 1999  
## 2 audi a4 1999  
## 3 audi a4 2008  
## 4 audi a4 2008  
## 5 audi a4 1999  
## 6 audi a4 1999  
## 7 audi a4 2008  
## 8 audi a4 quattro 1999  
## 9 audi a4 quattro 1999  
## 10 audi a4 quattro 2008  
## # … with 224 more rows

view(diamonds)  
  
#20) ;]To the diamonds data table, apply R code to produce the table give below.   
#Note that the variable name color has been changed to Color.  
diamonds%>%  
 select(color)%>%  
 group\_by(color)%>%  
 count()

## # A tibble: 7 x 2  
## # Groups: color [7]  
## color n  
## <ord> <int>  
## 1 D 6775  
## 2 E 9797  
## 3 F 9542  
## 4 G 11292  
## 5 H 8304  
## 6 I 5422  
## 7 J 2808