Project

Ethan Hershman

2024-10-23

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)

cdi <- read.csv("/Users/boardman/Downloads/CDI.csv")
```

```
cdi
```

```
##     income degree region
## 1    20786   22.3      W
## 2    21729   22.8     NC
## 3    19517   25.4      S
## 4    19588   25.3      W
## 5    24400   27.8      W
## 6    16803   16.6     NE
## 7    18042   22.1      W
## 8    17461   13.7     NC
## 9    17823   18.8      S
## 10   21001   26.3      S
## 11   16721   15.2     NE
## 12   23779   32.8      W
## 13   25193   32.6      W
## 14   16399   14.9      W
## 15   21086   20.1     NC
## 16   25312   35.4     NE
## 17   20681   22.6     NE
## 18   24262   23.0     NE
## 19   31679   30.0     NE
## 20   22148   28.8      W
## 21   22355   18.8      S
```

```
## 22   15508   19.7      S
## 23   17185   14.6      W
## 24   18825   24.0      S
## 25   26884   30.2     NC
## 26   18934   23.0      W
## 27   23705   31.6     NC
## 28   24219   29.2     NC
## 29   18305   20.0     NE
## 30   19040   26.6     NC
## 31   18431   19.3     NC
## 32   33330   35.3     NE
## 33   20580   23.7     NC
## 34   26798   22.1      S
## 35   24875   25.8     NE
## 36   21610   18.5      S
## 37   21307   24.6      W
## 38   16876   20.2      S
## 39   32342   34.2     NE
## 40   18430   20.8      S
## 41   32230   31.7     NE
## 42   28999   49.0      S
## 43   22197   24.2     NE
## 44   25523   31.6      W
## 45   19148   21.4     NC
## 46   26772   36.0     NC
## 47   24523   24.0     NE
## 48   30081   49.9      S
## 49   18625   13.8      W
## 50   17263   15.5      S
## 51   19568   25.5      S
## 52   15399   23.8      W
## 53   28532   35.0      W
## 54   20924   13.5     NC
## 55   21641   26.3     NE
## 56   19895   22.2     NE
## 57   23470   25.0      S
## 58   28462   32.1     NE
## 59   17879   21.2      S
## 60   17662   18.4      S
## 61   24896   26.5     NE
## 62   22834   25.9     NE
## 63   21420   23.0      W
## 64   16365   16.9      W
## 65   15191   23.3      W
## 66   19140   19.3      S
## 67   23150   27.7     NE
## 68   18624   19.9      S
## 69   28819   31.3      W
## 70   22819   31.6      S
## 71   18611   20.0     NC
## 72   26909   34.4     NE
## 73   23603   33.3      S
## 74   17741   22.6      S
## 75   17866   18.3     NE
```

```
## 76   11545   15.2      S
## 77   16194   17.5      W
## 78   19215   23.7      W
## 79   18340   34.7      S
## 80   18410   20.0     NC
## 81   27391   28.4     NE
## 82   18463   19.7     NE
## 83   23658   24.8     NE
## 84   21005   32.7      S
## 85   15881   13.3      W
## 86   22548   24.8     NE
## 87   27378   32.0     NC
## 88   18583   19.7     NC
## 89   20942   28.3      S
## 90   19505   24.4      S
## 91   18521   15.9     NE
## 92   19295   23.7      S
## 93   19930   21.0     NE
## 94   18674   20.7     NC
## 95   16578   22.4      S
## 96   26248   25.0     NE
## 97   20303   28.8     NC
## 98   15453   13.2      W
## 99   17518   26.7      W
## 100  16327   12.8     NC
## 101  19401   24.4     NE
## 102  22156   29.0      W
## 103  18545   19.3      W
## 104  17815   17.0     NC
## 105  19073   17.6     NE
## 106  21973   18.7     NE
## 107  17101   18.8      S
## 108  21933   33.0      S
## 109  22284   25.2     NE
## 110  20997   30.7      W
## 111  21500   22.2     NE
## 112  20974   15.3     NE
## 113  16829   12.8     NC
## 114  22797   24.6      S
## 115  20658   35.3      S
## 116  18878   16.7     NE
## 117  31520   36.7     NE
## 118  19629   24.9     NC
## 119  14835   12.9      S
## 120  19276   22.2     NC
## 121  17668   20.4      S
## 122  16807   25.8      W
## 123  18113   15.3     NC
## 124  23008   23.6     NE
## 125  17697   25.5      S
## 126  22507   35.2      W
## 127  22055   24.5      W
## 128   8899   11.5      S
## 129  17881   27.5      S
```

```
## 130  14389   15.5      S
## 131  24732   34.7     NE
## 132  15648   14.8      S
## 133  15238   13.0      W
## 134  17069   15.4     NE
## 135  21902   26.6      W
## 136  16898   14.3     NC
## 137  20087   34.2     NC
## 138  16365   20.6      W
## 139  18787   18.0     NC
## 140  19465   21.5      W
## 141  26156   40.5     NC
## 142  19861   29.6      S
## 143  18225   23.5      S
## 144  20349   24.8      S
## 145  17268   18.7      W
## 146  19502   13.9     NE
## 147  19655   15.1     NE
## 148  22581   26.4     NE
## 149  17382   23.9      S
## 150  18877   16.4      S
## 151  16405   13.1     NE
## 152  26026   29.5     NE
## 153  17874   21.0      S
## 154  21684   21.4     NC
## 155  14710   11.8      W
## 156  19932   29.8      W
## 157  19788   19.5     NE
## 158  23004   27.1     NC
## 159  19123   19.0     NC
## 160  16015   22.4      S
## 161  21003   28.3     NE
## 162  16750   18.7     NC
## 163  15124   17.0      S
## 164  19785   19.6     NE
## 165  17885   26.3      S
## 166  17137   28.0      S
## 167  18242   19.7      S
## 168  22782   41.9     NC
## 169  15701   22.2      W
## 170  17458   29.2     NC
## 171  13944    9.1      S
## 172  19942   23.6      W
## 173  24948   21.9      S
## 174  16331   16.2     NE
## 175  21123   27.6     NC
## 176  12923   16.6      S
## 177  17801   32.3      S
## 178  16006   12.3     NC
## 179  20645   24.1      S
## 180  26757   33.0     NE
## 181  16116   13.0      W
## 182  16256   14.0     NC
## 183  22303   39.1      S
```

```
## 184  11467   26.2      W
## 185  16190   14.7     NC
## 186  15392   18.2      S
## 187  16412   16.8      S
## 188   9728   12.0      S
## 189  22173   24.8     NE
## 190  20259   21.8     NE
## 191  21327   20.7      W
## 192  16215   26.4      S
## 193  18376   16.7     NC
## 194  16477   16.7     NE
## 195  17980   14.4     NC
## 196  16337   18.2      S
## 197  18336   16.7     NE
## 198  17211   19.2     NC
## 199  21770   25.9     NE
## 200  21362   27.6     NE
## 201  33180   38.3     NE
## 202  17418   15.5      S
## 203  18990   30.1      S
## 204  16790   16.8      W
## 205  18348   18.6     NE
## 206  37541   44.0      W
## 207  18523   18.1     NE
## 208  22025   29.7      W
## 209  16022   17.5      W
## 210  16144   11.4     NC
## 211  15776   14.3      S
## 212  18301   30.2      S
## 213  19320   30.6      S
## 214  21421   42.1      W
## 215  24035   16.4     NE
## 216  18288   27.1     NC
## 217  15443   23.4      S
## 218  16647   13.6     NE
## 219  16963   14.8     NE
## 220  17744   19.3      S
## 221  17221   22.9      W
## 222  17776   18.6      S
## 223  20543   27.6      S
## 224  19692   17.5     NC
## 225  17816   27.6     NC
## 226  18753   21.2     NC
## 227  18058   20.7     NE
## 228  16904   13.0     NC
## 229  17997   15.5      S
## 230  17469   24.2      S
## 231  16630   20.7     NC
## 232  17192   17.6      S
## 233  18786   24.9      W
## 234  16625   13.6     NE
## 235  15419   17.3     NC
## 236  19254   22.9     NE
## 237  13802   11.5      S
```

```
## 238  18490   17.7     NC
## 239  16422   37.1      S
## 240  17951   15.1      S
## 241  13536   17.2      S
## 242  17009   19.8      W
## 243  15941   17.3      S
## 244  14925   16.6      S
## 245  15374   13.7      W
## 246  13394   23.5      W
## 247  18360   18.7     NC
## 248  27546   46.9      S
## 249  23267   28.1     NE
## 250  17140   32.3      W
## 251  15162   11.9     NE
## 252  21855   21.0     NC
## 253  18342   19.5     NC
## 254  17084   19.4      S
## 255  20941   21.5      S
## 256  15051   19.5      W
## 257  16171   14.7      S
## 258  19238   33.4      S
## 259  16058   34.6      S
## 260  18857   25.2     NE
## 261  15505   16.6      S
## 262  13961   12.0      W
## 263  19601   22.4     NC
## 264  16319   10.8      S
## 265  18426   16.5     NC
## 266  16934   19.1      S
## 267  14443   25.9      S
## 268  25161   25.0     NE
## 269  16957   34.1     NC
## 270  20168   22.7      W
## 271  15896    9.0     NC
## 272  30242   52.3      S
## 273  15327   18.4      S
## 274  14968   14.7      S
## 275  18126   21.0      S
## 276  13691   16.3      S
## 277  18824   21.6     NE
## 278  18093   16.0     NC
## 279  16868   22.5      S
## 280  17908   19.0     NE
## 281  14473   10.8     NE
## 282  14134   10.3     NC
## 283  16232   16.7     NC
## 284  17312   24.7      W
## 285  20086   20.5     NE
## 286  19558   22.3     NC
## 287  14767   11.1     NC
## 288  15301   18.0      W
## 289  16770   14.2     NC
## 290  17774   19.5     NE
## 291  18395   20.0     NC
```

```
## 292  15853    8.1     NE
## 293  17496   12.7      S
## 294  25589   22.3      S
## 295  17251   15.6     NE
## 296  16924   16.9      S
## 297  17511   19.8      S
## 298  15113   20.0      S
## 299  19954   22.0      S
## 300  16231   14.5     NC
## 301  14137   13.1      S
## 302  17548   17.0      S
## 303  10190   13.4      S
## 304  15750   12.9     NC
## 305  20679   23.0     NE
## 306  17818   15.0     NC
## 307  16676   12.2     NC
## 308  16277   13.7      W
## 309  15521   17.7     NE
## 310  17853   31.9     NE
## 311  15582   17.6      W
## 312  20682   26.2     NC
## 313  17480   10.7     NC
## 314  14051    9.3     NE
## 315  14205   12.9      S
## 316  17129   23.1      S
## 317  14693   16.0      S
## 318  15803   21.0      S
## 319  15747   15.6      S
## 320  24132   28.2     NE
## 321  16031   17.6     NC
## 322  13869   18.9      S
## 323  16935   17.0      S
## 324  15197   14.2     NE
## 325  19727   30.3      W
## 326  17182   16.7     NC
## 327  17645   18.2     NC
## 328  14934   24.6      S
## 329  16742   13.3     NC
## 330  20068   20.9     NE
## 331  16819   10.8     NE
## 332  18161   26.0     NC
## 333  15944   13.8     NC
## 334  11379   21.9      W
## 335  14743   19.1      S
## 336  17278   10.5     NC
## 337   8973   11.1      S
## 338  15874   18.4      W
## 339  19940   34.0     NE
## 340  14615   14.6      S
## 341  16713   16.9      S
## 342  24405   24.9     NE
## 343  16018   11.7     NC
## 344  15847   30.7     NC
## 345  14779   10.5     NE
```

```
## 346  18961   29.0     NC
## 347  17566   18.5     NC
## 348  21944   29.2     NE
## 349  16412   13.0     NC
## 350  17338   12.7     NC
## 351  16002   22.0      W
## 352  14814   15.7      S
## 353  15079   10.0      S
## 354  16191   11.6     NC
## 355  19250   20.8      W
## 356  18526   21.3     NC
## 357  15476   32.3     NE
## 358  18008   13.6     NC
## 359  22002   19.6      S
## 360  14197   14.0      W
## 361  17119   16.5      S
## 362  18892   18.0     NC
## 363  12641   35.8      S
## 364  14834   12.9     NE
## 365  16281   12.4     NE
## 366  15177   13.6     NE
## 367  17898   17.2      S
## 368  16728   18.5      W
## 369  17119   21.2      S
## 370  20600   25.4     NE
## 371  15697   16.5     NC
## 372  16021   20.7      S
## 373  16138   20.0      W
## 374  14766   11.4      S
## 375  14757   17.5     NC
## 376  15778   12.3     NE
## 377  15501   18.7      S
## 378  17396   14.2      S
## 379  18021   14.8     NC
## 380  11396    8.2      W
## 381  13776   14.2      S
## 382  17131   18.1     NE
## 383  21153   19.6     NC
## 384  16305   13.5     NC
## 385  13475   14.4      S
## 386  14961   14.8      S
## 387  16500   11.8     NE
## 388  17272   21.5      W
## 389  14736   20.0      S
## 390  17522   21.9     NC
## 391  17332   23.3      W
## 392  17175   36.5     NC
## 393  12704   15.1     NE
## 394  16499   11.0     NC
## 395  13228   18.4      S
## 396  31699   48.5      S
## 397  14946   15.0     NE
## 398  16362   13.4      S
## 399  15205   13.6     NE
```

```
## 400  22668   22.3      W
## 401  15691   11.7      S
## 402  19449   29.1     NE
## 403  16542   11.4     NC
## 404  14523    9.7      W
## 405  14266   32.9     NC
## 406  25681   36.2     NC
## 407  12597    8.5     NC
## 408  17306   14.6      S
## 409  15852   21.9      S
## 410  30255   34.6     NE
## 411  16451   11.2      S
## 412  13681   17.7      W
## 413  16655   11.7     NC
## 414  16119   10.5     NC
## 415  11490   12.7      W
## 416  19345   26.4      S
## 417  14721    9.1      S
## 418  20515   29.5     NC
## 419  15036   18.1     NC
## 420  16029   17.9      S
## 421  16154   12.6     NE
## 422  10849   11.0      S
## 423  16775   17.7      S
## 424  13350   12.7      S
## 425  17182   21.7     NE
## 426  18061   13.8     NC
## 427  16342   15.5     NC
## 428  16514   14.4      S
## 429  16275   26.5     NC
## 430  11803   15.0      S
## 431  16137   25.4     NC
## 432  18070   16.8     NE
## 433  13907    9.0      W
## 434  16464   13.9     NC
## 435  19317   16.2      S
## 436  13919    9.7      S
## 437  27125   20.3      S
## 438  13169   16.5      S
## 439  18504   17.8      W
## 440  16458   15.5      S
```

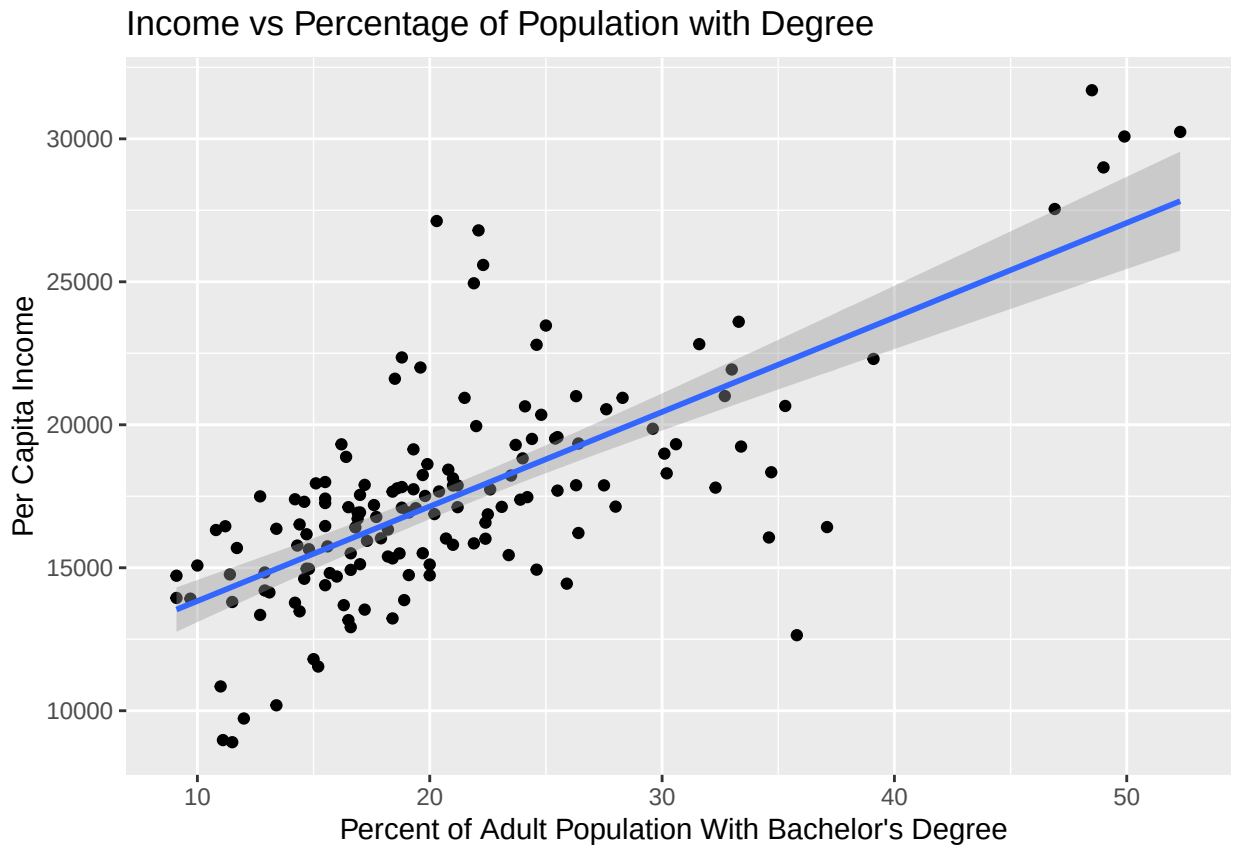
```
west <- cdi %>% filter(region == "W")
south <- cdi %>% filter(region == "S")
norcen <- cdi %>% filter(region == "NC")
noreast <- cdi %>% filter(region == "NE")
```

```
qplot(degree, income, data = south) + geom_smooth(method = 'lm') + labs(title =
"Income vs Percentage of Population with Degree", x = "Percent of Adult Population With Bachelor's Degr
y = "Per Capita Income")
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
```

```
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```




```
var(south$income)
```

```
## [1] 14775004
```

```
model2 = lm(income ~ degree, data = south)
model2
```

```
##
## Call:
## lm(formula = income ~ degree, data = south)
##
## Coefficients:
## (Intercept)       degree
##     10529.8        330.6
```

```
anova(model2)
```

```
## Analysis of Variance Table
```

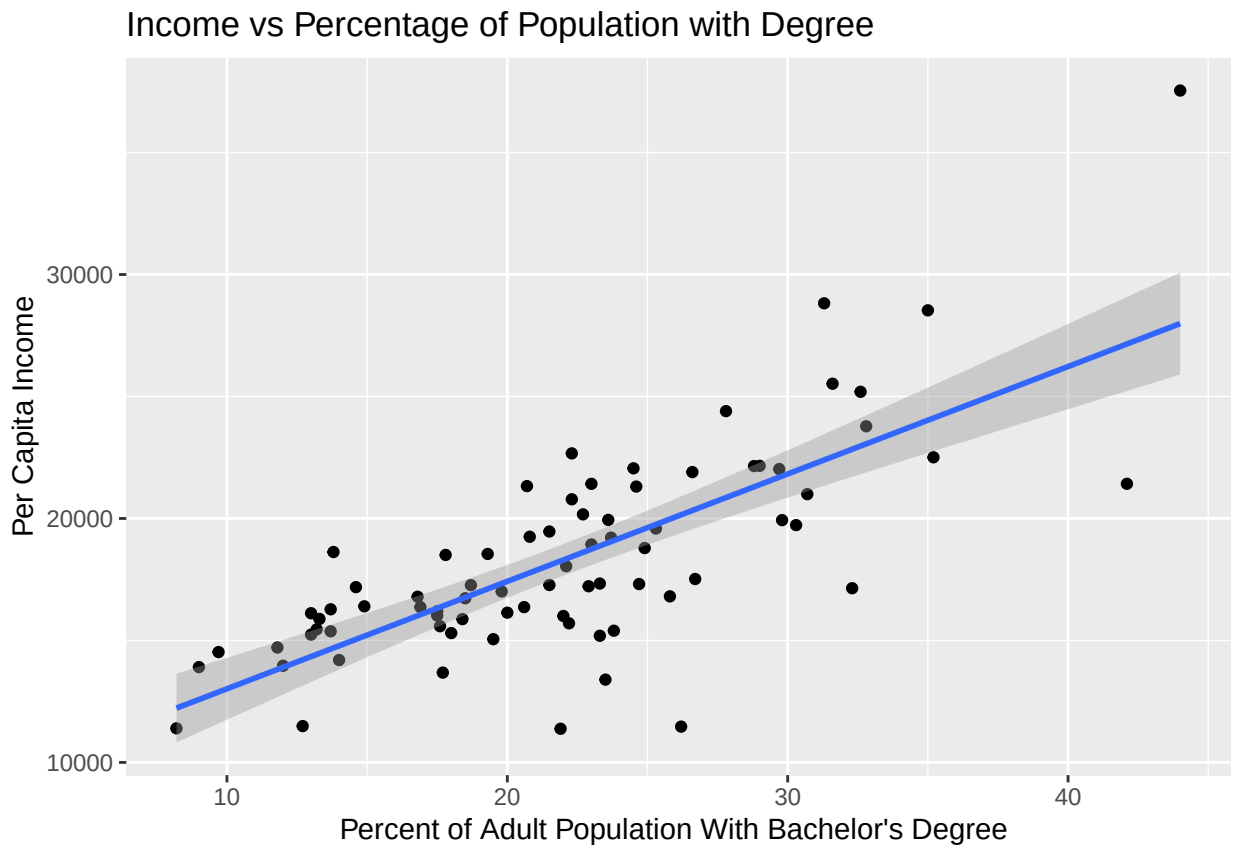
```
##
## Response: income
##             Df     Sum Sq    Mean Sq F value    Pr(>F)
## degree       1 1109873245 1109873245  148.49 < 2.2e-16 ***
## Residuals 150 1121152411    7474349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sse2 <- 19449165 * 50
ssto2 <- (813271618 + sse2)*51
1 - (sse2/ssto2)
```

```
## [1] 0.9893221
```

```
qplot(degree, income, data = west) + geom_smooth(method = 'lm') + labs(title =
"Income vs Percentage of Population with Degree", x = "Percent of Adult Population With Bachelor's Degr
y = "Per Capita Income")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
var(west$income)
```

```
## [1] 18287100
```

```r
model = lm(income ~ degree, data = west)
model
```

```
##
## Call:
## lm(formula = income ~ degree, data = west)
##
## Coefficients:
## (Intercept)       degree
##      8615.1        440.3
```

```r
anova(model)
```

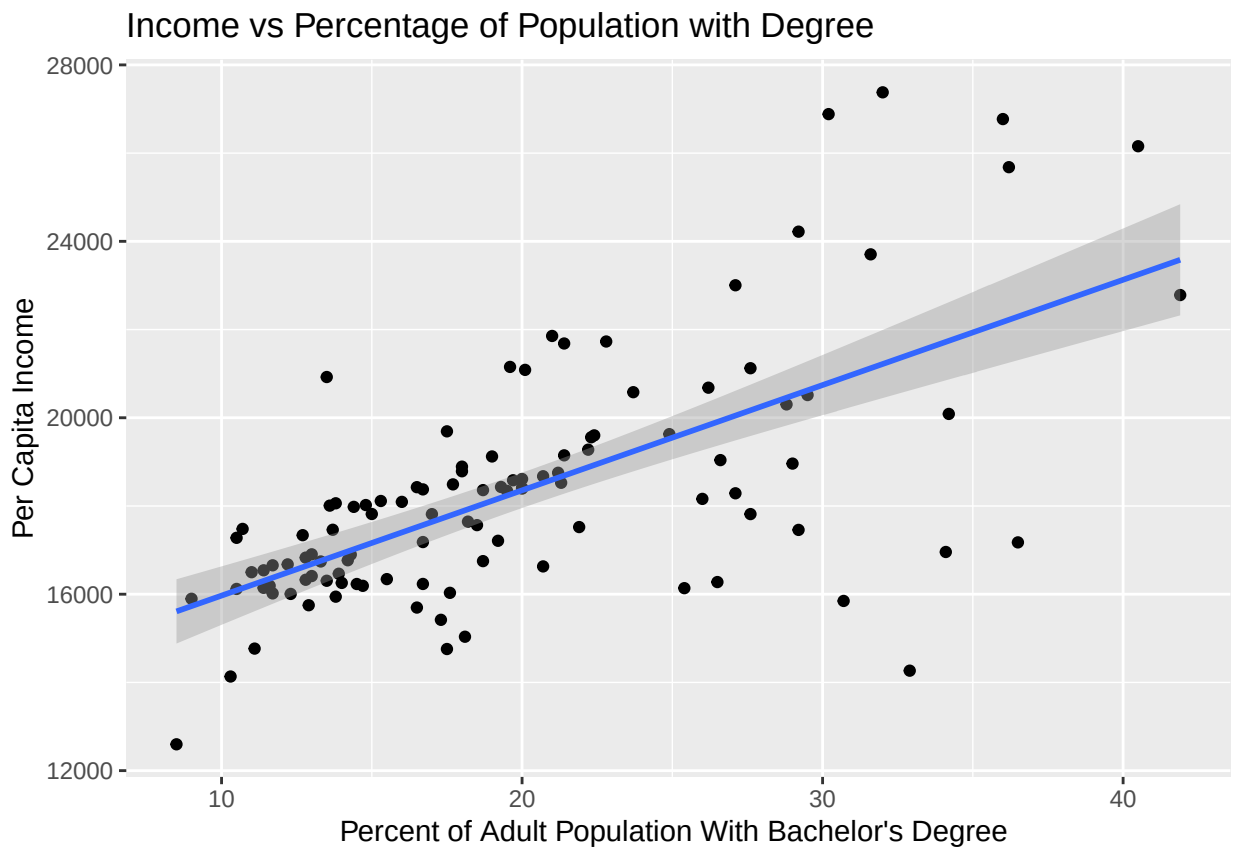
```
## Analysis of Variance Table
##
## Response: income
##           Df    Sum Sq   Mean Sq F value    Pr(>F)
## degree     1 773745787 773745787  94.195 6.856e-15 ***
## Residuals 75 616073841   8214318
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
sse <- 8214318 * 75
ssto <- (773745787 + sse)*76
1 - (sse/ssto)
```

```
## [1] 0.9941674
```

```r
qplot(degree, income, data = norcen) + geom_smooth(method = 'lm') + labs(title =
"Income vs Percentage of Population with Degree", x = "Percent of Adult Population With Bachelor's Degr
y = "Per Capita Income")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
var(norcen$income)
```

```
## [1] 7537475
```

```
model3 = lm(income ~ degree, data = norcen)
model3
```

```
##
## Call:
## lm(formula = income ~ degree, data = norcen)
##
## Coefficients:
## (Intercept)       degree
##     13581.4        238.7
```

```
anova(model3)
```

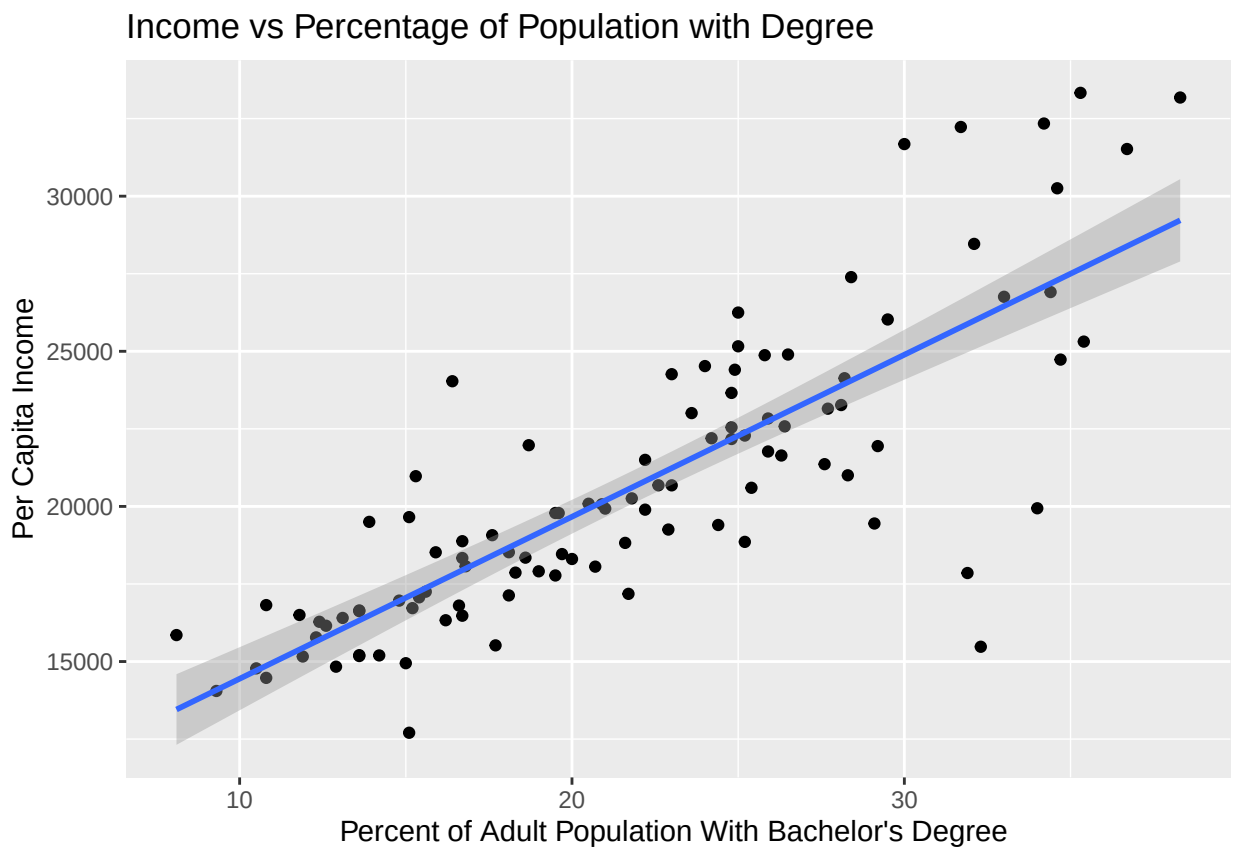
```
## Analysis of Variance Table
##
## Response: income
##            Df    Sum Sq   Mean Sq F value    Pr(>F)
## degree      1 338907694 338907694  76.826 3.344e-14 ***
## Residuals 106 467602149   4411341
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
sse3 <- 4411341 * 106
ssto3 <- (338907694 + sse3)*107
1 - (sse3/ssto3)
```

```
## [1] 0.9945815
```

```r
qplot(degree, income, data = noreast) + geom_smooth(method = 'lm') + labs(title =
"Income vs Percentage of Population with Degree", x = "Percent of Adult Population With Bachelor's Degr
y = "Per Capita Income")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
var(noreast$income)
```

```
## [1] 21483857
```

```r
model4 = lm(income ~ degree, data = noreast)
model4
```

```
##
## Call:
## lm(formula = income ~ degree, data = noreast)
##
```

```
## Coefficients:
## (Intercept)       degree
##      9223.8        522.2
```

```
anova(model4)
```

```
## Analysis of Variance Table
##
## Response: income
##             Df     Sum Sq    Mean Sq F value    Pr(>F)
## degree       1 1450517671 1450517671  197.75 < 2.2e-16 ***
## Residuals 101  740835765    7335008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sse4 <- 7335008 * 101
ssto4 <- (1450517671 + sse4)*102
1 - (sse4/ssto4)
```

```
## [1] 0.9966856
```