data151_EDA

Tyler Bontrager

2022-10-11

```r
# IMPORTING DATASETS
tuition_cost <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/d
```

```
## Rows: 2973 Columns: 10
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr (5): name, state, state_code, type, degree_length
## dbl (5): room_and_board, in_state_tuition, in_state_total, out_of_state_tuit...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
tc = tuition_cost

tuition_income <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/
```

```
## Rows: 209012 Columns: 7
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr (4): name, state, campus, income_lvl
## dbl (3): total_price, year, net_cost
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
ti = tuition_income

salary_potential <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/mast
```

```
## Rows: 935 Columns: 7
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr (2): name, state_name
## dbl (5): rank, early_career_pay, mid_career_pay, make_world_better_percent, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
sp = salary_potential

historical_tuition <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/mas
```

```
## Rows: 270 Columns: 4
## -- Column specification ------------------------------------------------------
## Delimiter: ","
```

```
## chr (3): type, year, tuition_type
## dbl (1): tuition_cost
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
ht = historical_tuition
```

```r
diversity_school <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/maste
```

```
## Rows: 50655 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): name, state, category
## dbl (2): total_enrollment, enrollment
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
ds = diversity_school
```

```r
head(tc)
```

```
## # A tibble: 6 x 10
##   name        state state~1 type  degre~2 room_~3 in_st~4 in_st~5 out_o~6 out_o~7
##   <chr>       <chr> <chr>   <chr> <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 Aaniiih N~ Mont~ MT      Publ~ 2 Year       NA    2380    2380    2380    2380
## 2 Abilene C~ Texas TX      Priv~ 4 Year    10350   34850   45200   34850   45200
## 3 Abraham B~ Geor~ GA      Publ~ 2 Year     8474    4128   12602   12550   21024
## 4 Academy C~ Minn~ MN      For ~ 2 Year       NA   17661   17661   17661   17661
## 5 Academy o~ Cali~ CA      For ~ 4 Year    16648   27810   44458   27810   44458
## 6 Adams Sta~ Colo~ CO      Publ~ 4 Year     8782    9440   18222   20456   29238
## # ... with abbreviated variable names 1: state_code, 2: degree_length,
## #   3: room_and_board, 4: in_state_tuition, 5: in_state_total,
## #   6: out_of_state_tuition, 7: out_of_state_total
```

```r
head(ti)
```

```
## # A tibble: 6 x 7
##   name                                state total_p~1  year campus net_c~2 incom~3
##   <chr>                               <chr>     <dbl> <dbl> <chr>    <dbl> <chr>
## 1 Piedmont International University   NC        20174  2016 On Ca~   11475 0 to 3~
## 2 Piedmont International University   NC        20174  2016 On Ca~   11451 30,001~
## 3 Piedmont International University   NC        20174  2016 On Ca~   16229 48_001~
## 4 Piedmont International University   NC        20174  2016 On Ca~   15592 75,001~
## 5 Piedmont International University   NC        20514  2017 On Ca~   11668. 0 to 3~
## 6 Piedmont International University   NC        20514  2017 On Ca~   11644. 30,001~
## # ... with abbreviated variable names 1: total_price, 2: net_cost,
## #   3: income_lvl
```

```r
head(sp)
```

```
## # A tibble: 6 x 7
##    rank name                            state~1 early~2 mid_c~3 make_~4 stem_~5
##   <dbl> <chr>                           <chr>     <dbl>   <dbl>   <dbl>   <dbl>
## 1     1 Auburn University               Alabama   54400  104500      51      31
## 2     2 University of Alabama in Huntsv~ Alabama   57500  103900      59      45
```

```
## 3       3 The University of Alabama          Alabama    52300   97400      50       15
## 4       4 Tuskegee University                Alabama    54500   93500      61       30
## 5       5 Samford University                 Alabama    48400   90500      52        3
## 6       6 Spring Hill College                Alabama    46600   89100      53       12
## # ... with abbreviated variable names 1: state_name, 2: early_career_pay,
## #   3: mid_career_pay, 4: make_world_better_percent, 5: stem_percent
```

```r
head(ht)
```

```
## # A tibble: 6 x 4
##   type             year    tuition_type    tuition_cost
##   <chr>            <chr>   <chr>                  <dbl>
## 1 All Institutions 1985-86 All Constant           10893
## 2 All Institutions 1985-86 4 Year Constant         12274
## 3 All Institutions 1985-86 2 Year Constant          7508
## 4 All Institutions 1985-86 All Current             4885
## 5 All Institutions 1985-86 4 Year Current          5504
## 6 All Institutions 1985-86 2 Year Current          3367
```

```r
head(ds)
```

```
## # A tibble: 6 x 5
##   name                        total_enrollment state   category        enrol~1
##   <chr>                                  <dbl> <chr>   <chr>             <dbl>
## 1 University of Phoenix-Arizona        195059 Arizona Women            134722
## 2 University of Phoenix-Arizona        195059 Arizona American India~     876
## 3 University of Phoenix-Arizona        195059 Arizona Asian              1959
## 4 University of Phoenix-Arizona        195059 Arizona Black             31455
## 5 University of Phoenix-Arizona        195059 Arizona Hispanic          13984
## 6 University of Phoenix-Arizona        195059 Arizona Native Hawaiia~    1019
## # ... with abbreviated variable name 1: enrollment
```

```r
str(tc)
```

```
## spec_tbl_df [2,973 x 10] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ name               : chr [1:2973] "Aaniiih Nakoda College" "Abilene Christian University" "Abraha
##  $ state              : chr [1:2973] "Montana" "Texas" "Georgia" "Minnesota" ...
##  $ state_code         : chr [1:2973] "MT" "TX" "GA" "MN" ...
##  $ type               : chr [1:2973] "Public" "Private" "Public" "For Profit" ...
##  $ degree_length      : chr [1:2973] "2 Year" "4 Year" "2 Year" "2 Year" ...
##  $ room_and_board     : num [1:2973] NA 10350 8474 NA 16648 ...
##  $ in_state_tuition   : num [1:2973] 2380 34850 4128 17661 27810 ...
##  $ in_state_total     : num [1:2973] 2380 45200 12602 17661 44458 ...
##  $ out_of_state_tuition: num [1:2973] 2380 34850 12550 17661 27810 ...
##  $ out_of_state_total : num [1:2973] 2380 45200 21024 17661 44458 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   name = col_character(),
##   ..   state = col_character(),
##   ..   state_code = col_character(),
##   ..   type = col_character(),
##   ..   degree_length = col_character(),
##   ..   room_and_board = col_double(),
##   ..   in_state_tuition = col_double(),
##   ..   in_state_total = col_double(),
##   ..   out_of_state_tuition = col_double(),
```

```
##    ..    out_of_state_total = col_double()
##    .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(ti)
```

```
## spec_tbl_df [209,012 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ name       : chr [1:209012] "Piedmont International University" "Piedmont International University"
## $ state      : chr [1:209012] "NC" "NC" "NC" "NC" ...
## $ total_price: num [1:209012] 20174 20174 20174 20174 20514 ...
## $ year       : num [1:209012] 2016 2016 2016 2016 2017 ...
## $ campus     : chr [1:209012] "On Campus" "On Campus" "On Campus" "On Campus" ...
## $ net_cost   : num [1:209012] 11475 11451 16229 15592 11668 ...
## $ income_lvl : chr [1:209012] "0 to 30,000" "30,001 to 48,000" "48_001 to 75,000" "75,001 to 110,000
##  - attr(*, "spec")=
##   .. cols(
##   ..    name = col_character(),
##   ..    state = col_character(),
##   ..    total_price = col_double(),
##   ..    year = col_double(),
##   ..    campus = col_character(),
##   ..    net_cost = col_double(),
##   ..    income_lvl = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(sp)
```

```
## spec_tbl_df [935 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ rank                    : num [1:935] 1 2 3 4 5 6 7 8 9 10 ...
## $ name                    : chr [1:935] "Auburn University" "University of Alabama in Huntsville" "
## $ state_name              : chr [1:935] "Alabama" "Alabama" "Alabama" "Alabama" ...
## $ early_career_pay        : num [1:935] 54400 57500 52300 54500 48400 46600 49100 48600 47700 48700
## $ mid_career_pay          : num [1:935] 104500 103900 97400 93500 90500 ...
## $ make_world_better_percent: num [1:935] 51 59 50 61 52 53 48 57 56 58 ...
## $ stem_percent            : num [1:935] 31 45 15 30 3 12 27 17 17 20 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..    rank = col_double(),
##   ..    name = col_character(),
##   ..    state_name = col_character(),
##   ..    early_career_pay = col_double(),
##   ..    mid_career_pay = col_double(),
##   ..    make_world_better_percent = col_double(),
##   ..    stem_percent = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(ht)
```

```
## spec_tbl_df [270 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ type        : chr [1:270] "All Institutions" "All Institutions" "All Institutions" "All Institutio
## $ year        : chr [1:270] "1985-86" "1985-86" "1985-86" "1985-86" ...
## $ tuition_type: chr [1:270] "All Constant" "4 Year Constant" "2 Year Constant" "All Current" ...
## $ tuition_cost: num [1:270] 10893 12274 7508 4885 5504 ...
##  - attr(*, "spec")=
```

```
##    .. cols(
##    ..   type = col_character(),
##    ..   year = col_character(),
##    ..   tuition_type = col_character(),
##    ..   tuition_cost = col_double()
##    .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(ds)
```

```
## spec_tbl_df [50,655 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ name            : chr [1:50655] "University of Phoenix-Arizona" "University of Phoenix-Arizona" "U
##  $ total_enrollment: num [1:50655] 195059 195059 195059 195059 195059 ...
##  $ state           : chr [1:50655] "Arizona" "Arizona" "Arizona" "Arizona" ...
##  $ category        : chr [1:50655] "Women" "American Indian / Alaska Native" "Asian" "Black" ...
##  $ enrollment      : num [1:50655] 134722 876 1959 31455 13984 ...
##  - attr(*, "spec")=
##    .. cols(
##    ..   name = col_character(),
##    ..   total_enrollment = col_double(),
##    ..   state = col_character(),
##    ..   category = col_character(),
##    ..   enrollment = col_double()
##    .. )
##  - attr(*, "problems")=<externalptr>
```

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## v purrr   0.3.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
?ggplot2
```

```
# Time to explore the data!

# Commenting out ggplot stuff to do dplyr first
#ggplot(ti,aes(x=year, y=total_price)) + geom_point()

#this is the median cost of attendance for instate/outstate

median_IN_COA <- tc %>%
  filter(degree_length=='4 Year') %>%
  group_by(state_code)%>%
  summarize(median_instate_COA = median(in_state_total))


median_IN_COA
```

```
## # A tibble: 53 x 2
##    state_code median_instate_COA
##    <chr>                   <dbl>
```

```
##  1 AK                    17017
##  2 AL                    18646
##  3 AR                    19023
##  4 AZ                    25037
##  5 CA                    30416
##  6 CO                    20976.
##  7 CT                    46455
##  8 DC                    50702.
##  9 DE                    26542
## 10 FL                    23352
## # ... with 43 more rows
```

```r
median_OUT_COA <- tc %>%
  filter(degree_length=='4 Year') %>%
  group_by(state_code)%>%
  summarize(median_outstate_COA = median(out_of_state_total))

median_OUT_COA
```

```
## # A tibble: 53 x 2
##     state_code median_outstate_COA
##     <chr>                    <dbl>
##  1 AK                       28604
##  2 AL                       27880
##  3 AR                       23709
##  4 AZ                       37190
##  5 CA                       36103
##  6 CO                       36096
##  7 CT                       48656.
##  8 DC                       50702.
##  9 DE                       30700
## 10 FL                       32000
## # ... with 43 more rows
```

```r
#  mutate(mean_instate_COA=mean(in_state_tuition)) %>%
#  mutate(mean_outofstate_COA=mean(out_of_state_tuition))%>%
```

```r
# library(tidyverse)
# library(janitor)
# library(glue)
#
# raw_df <- read_csv(here::here("2020", "2020-03-10", "all_schools.csv")) %>%
#   janitor::clean_names()
#
# raw_cost <- raw_df %>%
#   select(name = institution_name, state = state_abbreviation_hd2017,
#          contains("total_price_for_in_state_students")) %>%
#   pivot_longer(names_to = "category", values_to = "total_price", total_price_for_in_state_students_li
#   mutate(year = as.double(str_sub(category, -4)),
#          category = str_remove(category, "total_price_for_in_state_students_living_"),
#          campus = if_else(str_detect(category, "on_campus"), "On Campus", "Off Campus")) %>%
#   filter(!is.na(total_price)) %>%
#   select(-category)
#
# net_cost <- raw_df %>%
```

```r
#   select(name = institution_name, state = state_abbreviation_hd2017,
#          starts_with("average_net_price_income")) %>%
#   pivot_longer(names_to = "category", values_to = "net_cost",
#                average_net_price_income_0_30_000_students_awarded_title_iv_federal_financial_aid_2016
#   filter(!is.na(net_cost)) %>%
#   mutate(year = str_sub(category, -4, -3),
#          year = glue::glue("20{year}"),
#          year = as.double(year),
#          category = str_remove(category, "average_net_price_income_"),
#          category = str_remove(category, str_sub(category, -16)),
#          income_lvl = str_remove(category, "_students_awarded_title_iv_federal_financial_aid"),
#          income_lvl = case_when(
#            str_detect(income_lvl, "30_000") ~ "0 to 30,000",
#            str_detect(income_lvl, "30_001") ~ "30,001 to 48,000",
#            str_detect(income_lvl, "48_001") ~ "48_001 to 75,000",
#            str_detect(income_lvl, "75_001") ~ "75,001 to 110,000",
#            str_detect(income_lvl, "110_000") ~ "Over 110,000",
#            TRUE ~ NA_character_)
#          ) %>%
#   select(-category)
#
# full_dataset <- left_join(raw_cost, net_cost, by = c("name", "year", "state")) %>%
#   filter(!is.na(net_cost))
#
# full_dataset %>%
#   write_csv(here::here("2020", "2020-03-10", "tuition_income.csv"))

#code found on tidytuesday github by user jthomasmock https://github.com/jthomasmock
```

```r
library(tidyverse)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```r
library(glue)
```

```r
# raw_df <- read_csv(here::here("2020", "2020-03-10", "all_schools.csv")) %>%
#   janitor::clean_names()
#
# raw_cost <- raw_df %>%
#   select(name = institution_name, state = state_abbreviation_hd2017,
#          contains("total_price_for_in_state_students")) %>%
#   pivot_longer(names_to = "category", values_to = "total_price", total_price_for_in_state_students_li
#   mutate(year = as.double(str_sub(category, -4)),
#          category = str_remove(category, "total_price_for_in_state_students_living_"),
#          campus = if_else(str_detect(category, "on_campus"), "On Campus", "Off Campus")) %>%
#   filter(!is.na(total_price)) %>%
#   select(-category)
#
# net_cost <- raw_df %>%
```

```
#   select(name = institution_name, state = state_abbreviation_hd2017,
#          starts_with("average_net_price_income")) %>%
#   pivot_longer(names_to = "category", values_to = "net_cost",
#                average_net_price_income_0_30_000_students_awarded_title_iv_federal_financial_aid_2016
#   filter(!is.na(net_cost)) %>%
#   mutate(year = str_sub(category, -4, -3),
#          year = glue::glue("20{year}"),
#          year = as.double(year),
#          category = str_remove(category, "average_net_price_income_"),
#          category = str_remove(category, str_sub(category, -16)),
#          income_lvl = str_remove(category, "_students_awarded_title_iv_federal_financial_aid"),
#          income_lvl = case_when(
#            str_detect(income_lvl, "30_000") ~ "0 to 30,000",
#            str_detect(income_lvl, "30_001") ~ "30,001 to 48,000",
#            str_detect(income_lvl, "48_001") ~ "48,001 to 75,000",
#            str_detect(income_lvl, "75_001") ~ "75,001 to 110,000",
#            str_detect(income_lvl, "110_000") ~ "Over 110,000",
#            TRUE ~ NA_character_)
#          ) %>%
#   select(-category)
#
# full_dataset <- left_join(raw_cost, net_cost, by = c("name", "year", "state")) %>%
#   filter(!is.na(net_cost))
#
# full_dataset %>%
#   write_csv(here::here("2020", "2020-03-10", "tuition_income.csv"))

#code by jthomasmock
```

```
library(tidyverse)
library(rvest)
```

```
##
## Attaching package: 'rvest'

## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```
# scrape_salary <- function(state_name){
#
#   message(glue::glue("Scraping {state_name}!"))
#
#   Sys.sleep(5)
#
#   url <- glue::glue("https://www.payscale.com/college-salary-report/best-schools-by-state/bachelors/{
#
#   raw_table <- url %>%
#     read_html() %>%
#     html_table(fill = TRUE) %>%
#     .[[1]]
#
#
#   raw_table %>%
#     set_names(nm = c("rank", "name", "type", "early_career_pay", "mid_career_pay",
```

```
#                     "make_world_better_percent", "stem_percent")) %>%
#     as_tibble() %>%
#     mutate(rank = str_remove(rank, "Rank:"),
#            name = str_remove(name, "School Name:"),
#            type = str_remove(type, "School Type:"),
#            early_career_pay = parse_number(early_career_pay),
#            mid_career_pay = str_remove(mid_career_pay, "Mid-Career Pay:"),
#            mid_career_pay = parse_number(mid_career_pay),
#            make_world_better_percent = parse_number(make_world_better_percent),
#            stem_percent = parse_number(stem_percent),
#            state_name = str_to_title(state_name)
#     ) %>%
#     select(rank, name, state_name, everything(), -type)
# }
#
# all_states <- tolower(state.name) %>%
#   str_replace(" ", "-") %>%
#   map(scrape_salary) %>%
#   bind_rows() %>%
#   mutate(name = str_replace(name, "A & M", "A&M"))
#
# all_states %>%
#   write_csv(here::here("2020", "2020-03-10", "salary_potential.csv"))
```

```
#In the following graph, I want to find out whether colleges with higher STEM enrollment tend to cost m
```

```
jointisp = ti %>%
  left_join(sp) %>%
  group_by(stem_percent)%>%
  summarize(medianNet=median(net_cost))
```

```
## Joining, by = "name"
```

```
jointisp
```

```
## # A tibble: 75 x 2
##     stem_percent medianNet
##            <dbl>     <dbl>
##  1             0    21516.
##  2             1    19981.
##  3             2    21653
##  4             3    18443.
##  5             4    15509
##  6             5    17175
##  7             6    17907
##  8             7    19088
##  9             8    17178
## 10             9    17915
## # ... with 65 more rows
```

```
ggplot(jointisp, aes(stem_percent,medianNet)) + geom_point()
```
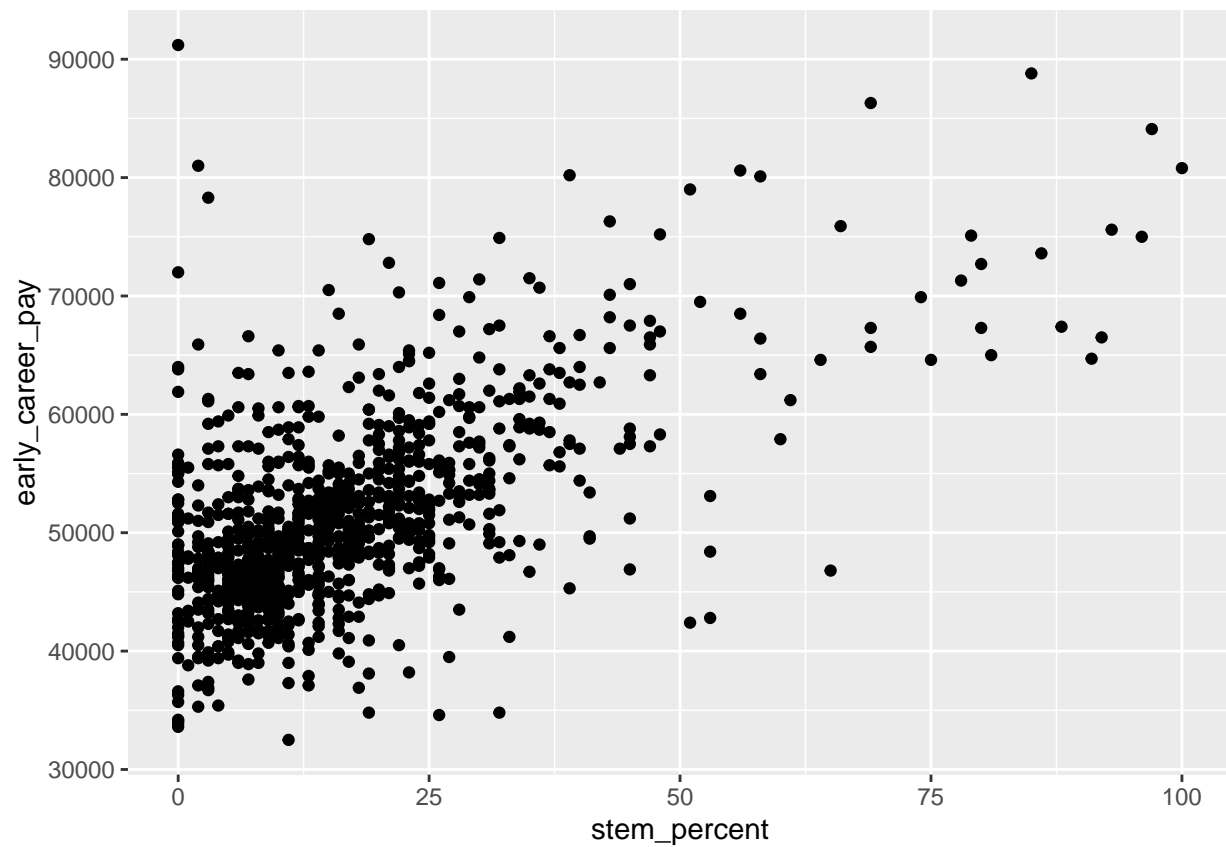
```
## Warning: Removed 1 rows containing missing values (geom_point).
```
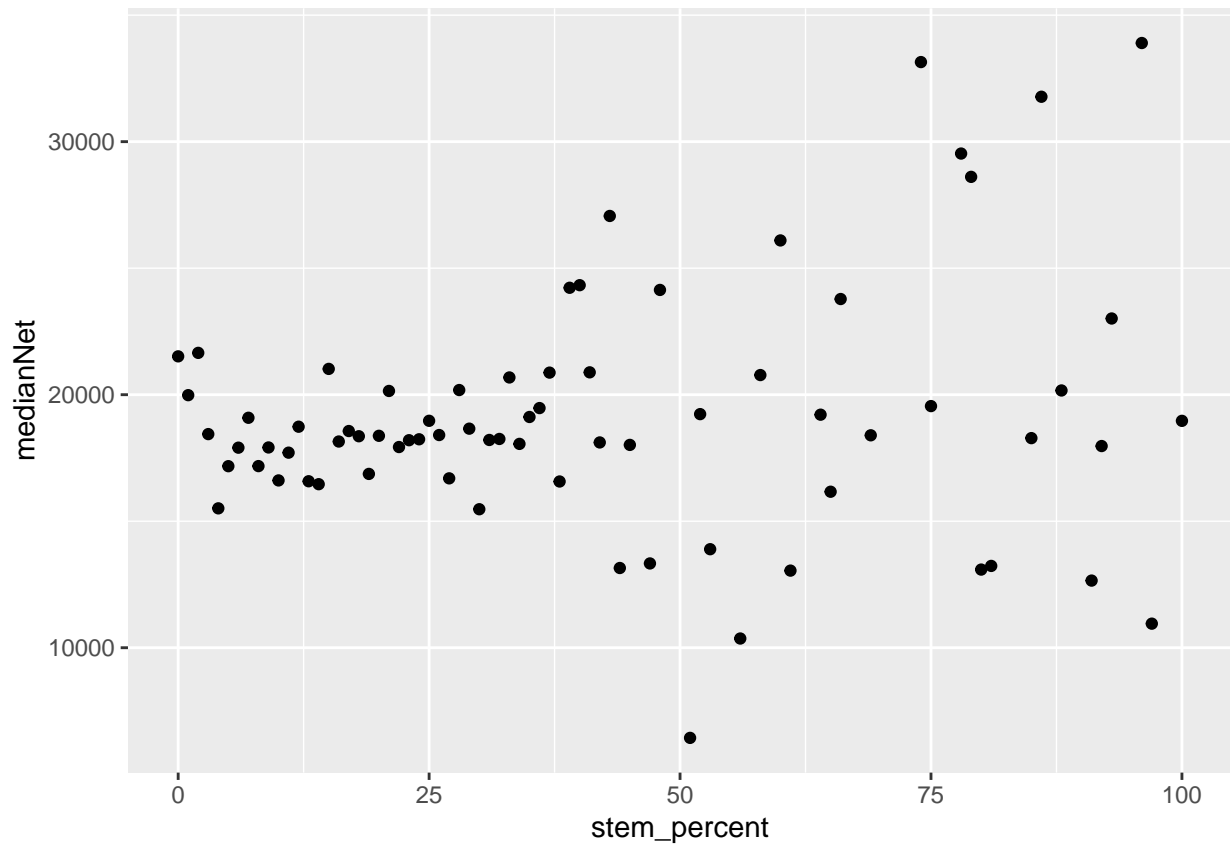
```
#My conclusion is that there is no association between higher STEM enrollment and median net cost.

#In this graph I want to see whether a higher STEM enrollment has a high association with early career p
# I will want to see whether the trend keeps for mid-career pay.

ggplot(sp, aes(stem_percent,early_career_pay)) + geom_point()
```
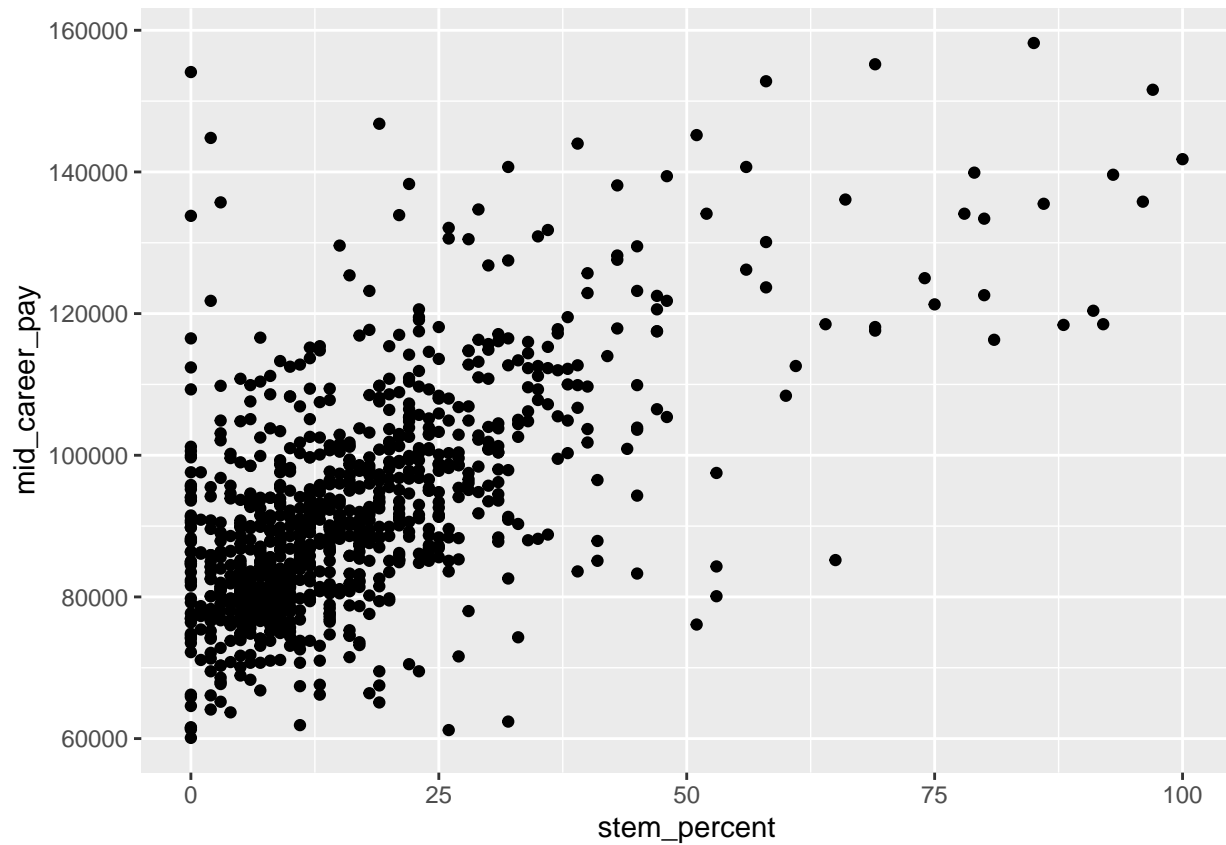
```
ggplot(sp, aes(stem_percent,mid_career_pay)) + geom_point()
```

```
#attempting to make the graphics easier to differentiate. I'm not sure what the difference is.

ggplot(sp, aes(x=stem_percent,y=early_career_pay)) + geom_point() + expand_limits(x=0,y=0)
```
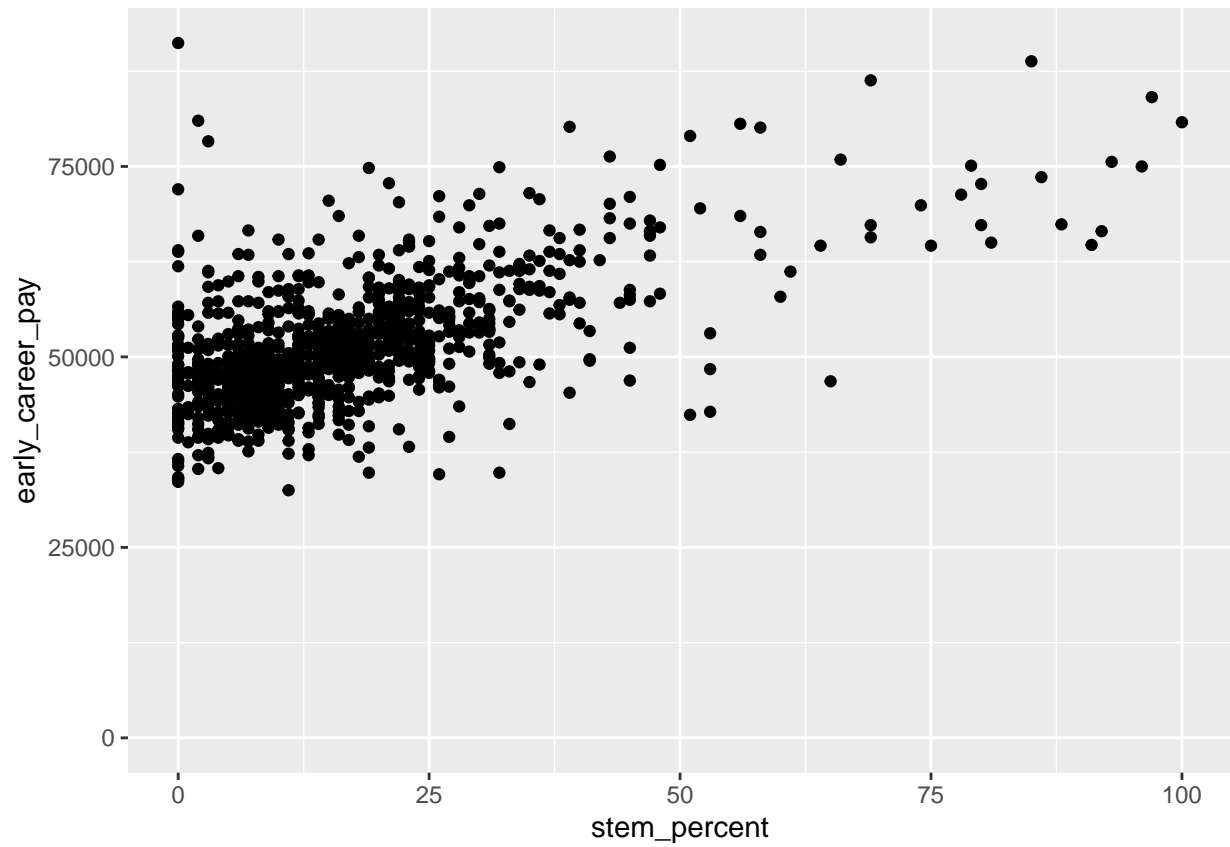
```
ggplot(sp, aes(x=stem_percent,y=mid_career_pay)) + geom_point() + expand_limits(x=0,y=0)
```
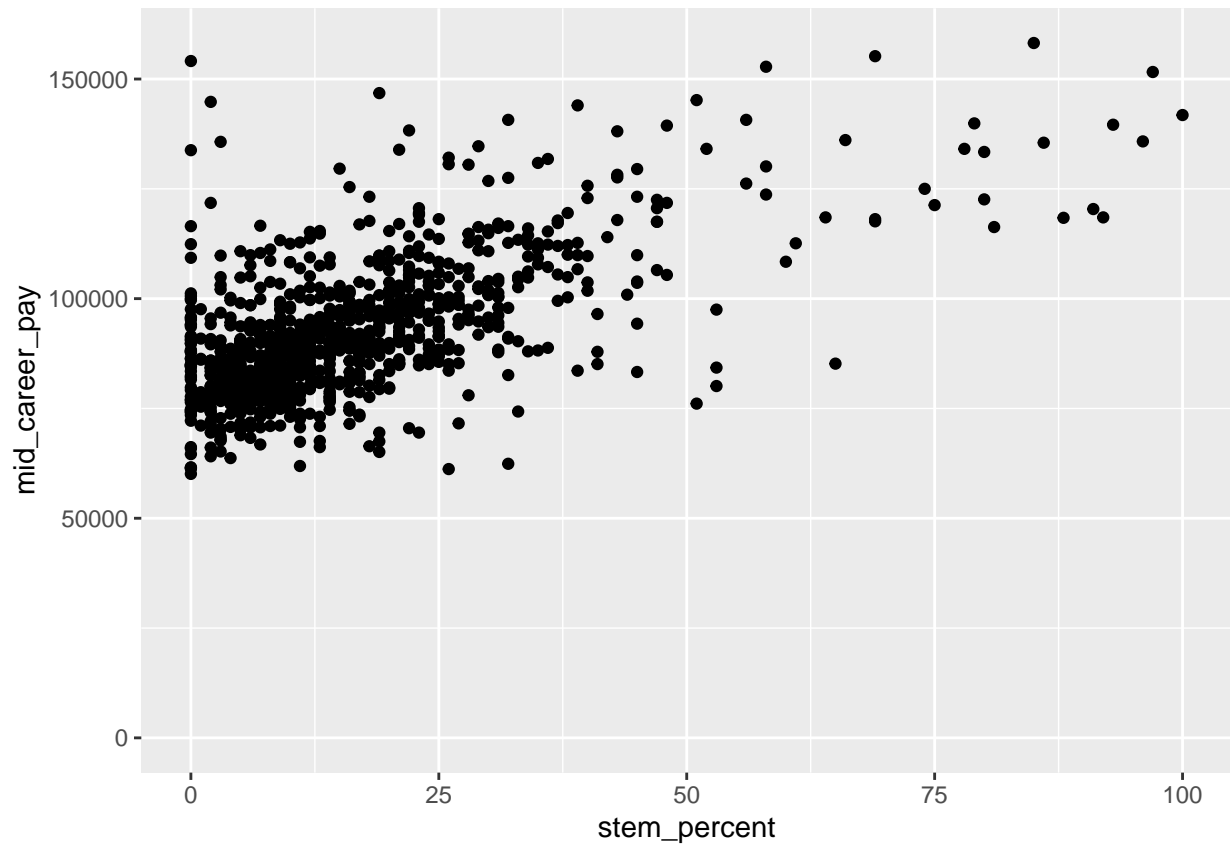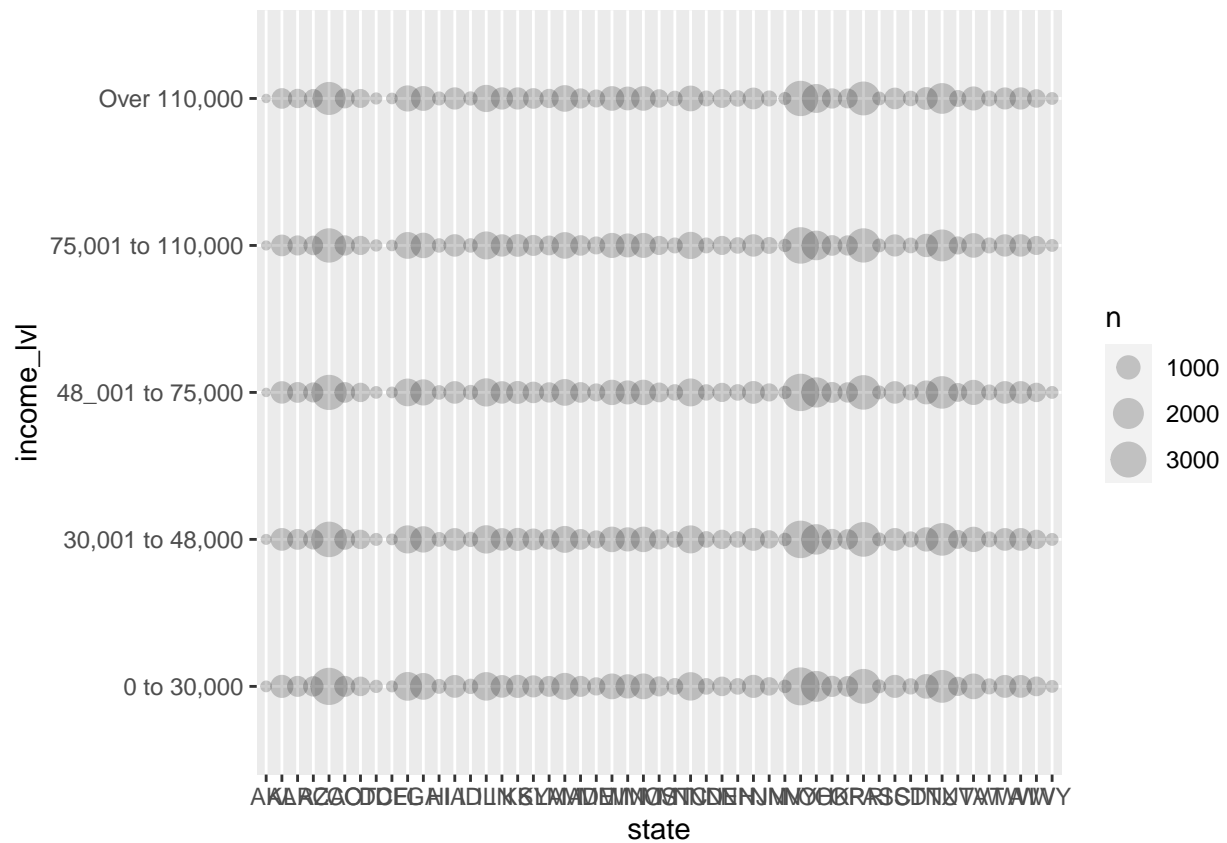
```
ggplot(sp, aes(x=stem_percent,y=mid_career_pay)) + geom_point() + expand_limits(x=0,y=0)
```
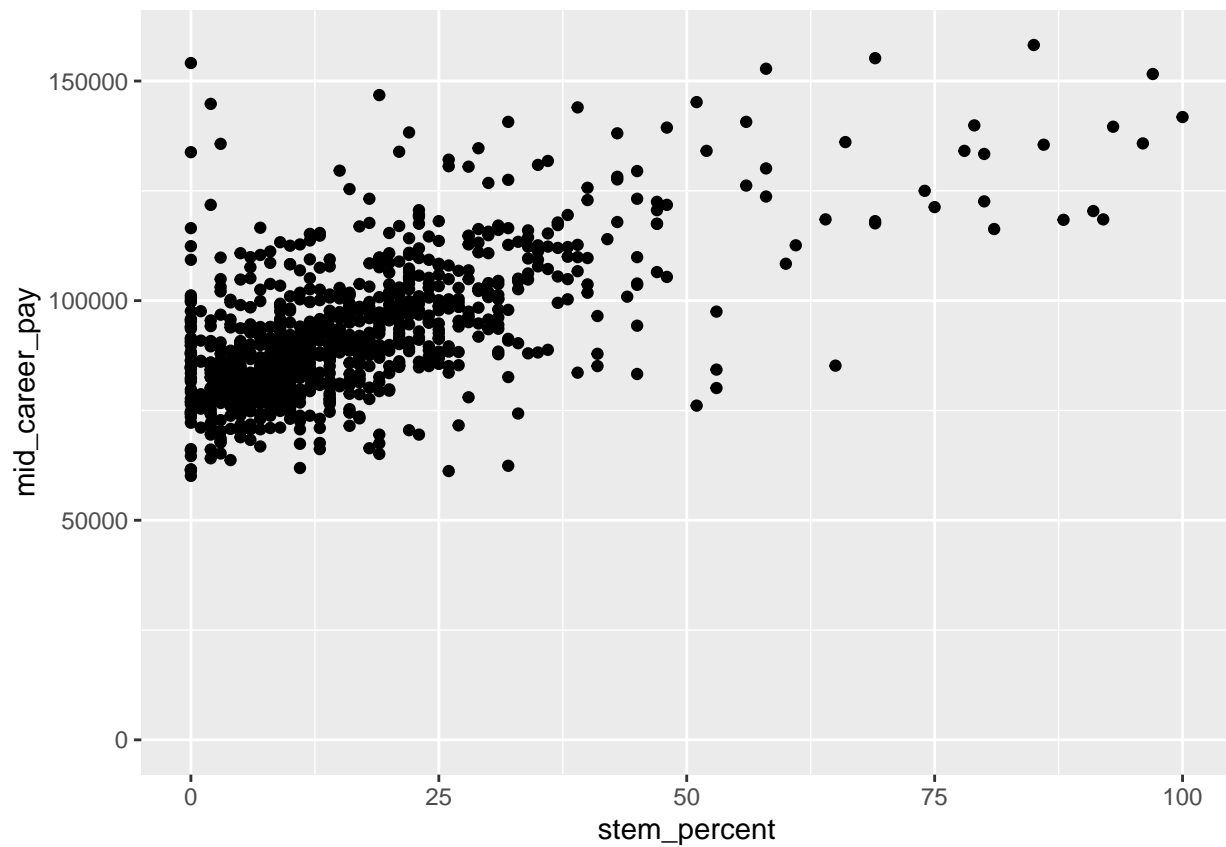
```
#incomeLvlByState = ti %>%


#ggplot(ti, aes(x=state, y=net_cost, fill=income_lvl))+
#  geom_bar(stat = "identity",
#          position="dodge")

ggplot(ti, aes(x=state,y=income_lvl)) + geom_count(alpha=0.2)
```
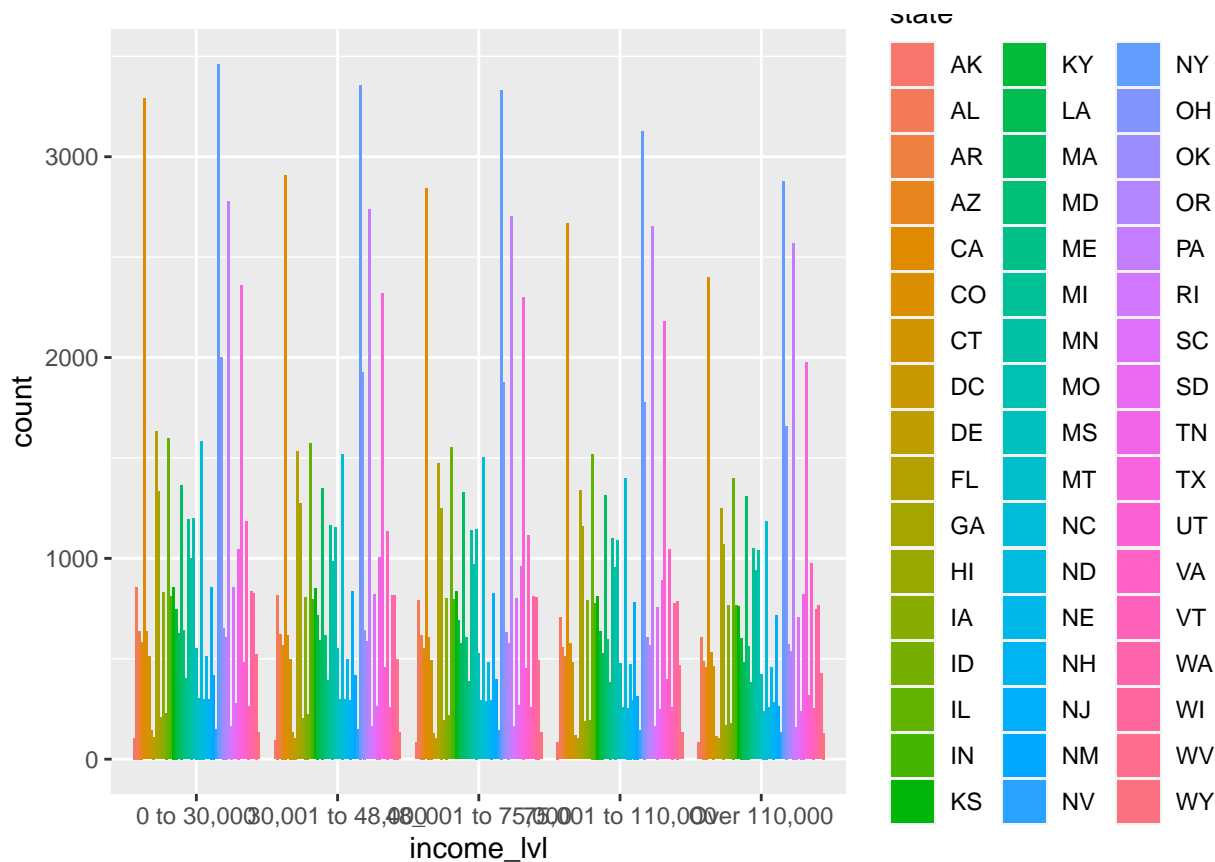
```
#factor(ti$income_lvl,levels='0 to 30,000', '30,001 to 48,000', '48_001 to 75,000', '75,001 to 110,000'
#incomeByState = ti %>%
#   group_by(income_lvl) %>%
#   summarize(state)
#incomeByState

ggplot(ti, aes(x=income_lvl, fill=state)) + geom_bar(position="dodge")
```
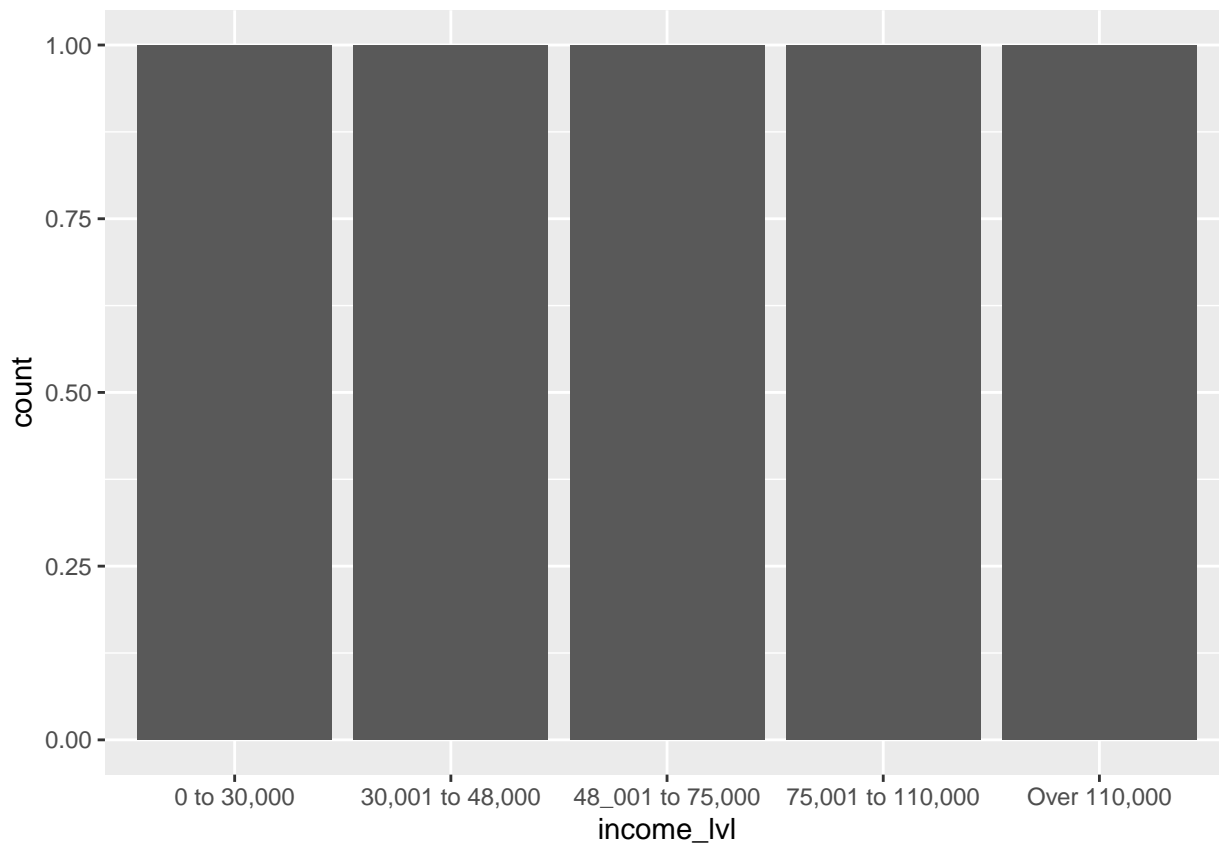
```r
income = ti %>%
  group_by(income_lvl) %>%
  summarize(Count=n()) %>%
  mutate(Percent = round((Count/sum(Count)*100))) %>%
  arrange(desc(Count))
income
```

```
## # A tibble: 5 x 3
##   income_lvl       Count Percent
##   <chr>            <int>   <dbl>
## 1 0 to 30,000      44969      22
## 2 30,001 to 48,000 43384      21
## 3 48_001 to 75,000 42600      20
## 4 75,001 to 110,000 40403     19
## 5 Over 110,000     37656      18
```

```r
ggplot(income, aes(x=income_lvl)) + geom_bar(position="fill")
```

```
incomeByState = ti %>%
  group_by(income_lvl,state) %>%
  summarize(Count=n()) %>%
  mutate(Percent = round((Count/sum(Count)*100))) %>%
  arrange(desc(Count))
```

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

incomeByState

```
## # A tibble: 255 x 4
## # Groups:   income_lvl [5]
##    income_lvl       state Count Percent
##    <chr>            <chr> <int>   <dbl>
##  1 0 to 30,000      NY     3460       8
##  2 30,001 to 48,000 NY     3357       8
##  3 48_001 to 75,000 NY     3333       8
##  4 0 to 30,000      CA     3290       7
##  5 75,001 to 110,000 NY    3125       8
##  6 30,001 to 48,000 CA     2909       7
##  7 Over 110,000     NY     2877       8
##  8 48_001 to 75,000 CA     2841       7
##  9 0 to 30,000      PA     2776       6
## 10 30,001 to 48,000 PA     2736       6
## # ... with 245 more rows
```

```
ggplot(incomeByState,aes(x=income_lvl,y=Count,fill=state)) + geom_col(stat="identity",pos="dodge")
```

```
## Warning: Ignoring unknown parameters: stat
```