Plots

2023-04-30

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- 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)
library(dplyr)

## load data
attorneys <- read.csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest 2023
attorney_time_entries <- read.csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/D
categories <- read.csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest 202
clients <- read.csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest 2023 D
questionposts <- read_csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest
```

```
## New names:
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
## * `` -> `...15`
## * `` -> `...16`
```

```
## * `` -> `...17`
## * `` -> `...18`
## * `` -> `...19`
## * `` -> `...20`
## * `` -> `...21`
## * `` -> `...22`
## * `` -> `...23`
## * `` -> `...24`
## * `` -> `...25`
## * `` -> `...26`
## * `` -> `...27`
## * `` -> `...28`
## * `` -> `...29`
## * `` -> `...30`
## * `` -> `...31`
## * `` -> `...32`
## * `` -> `...33`
## * `` -> `...34`
## * `` -> `...35`
## * `` -> `...36`
## * `` -> `...37`
## * `` -> `...38`
## * `` -> `...39`
## * `` -> `...40`
## * `` -> `...41`
## * `` -> `...42`
## * `` -> `...43`
## * `` -> `...44`
## * `` -> `...45`
## * `` -> `...46`
## * `` -> `...47`
## * `` -> `...48`
## * `` -> `...49`
## * `` -> `...50`
## * `` -> `...51`
## * `` -> `...52`
## * `` -> `...53`
## * `` -> `...54`
## * `` -> `...55`
## * `` -> `...56`
## * `` -> `...57`
## * `` -> `...58`
## * `` -> `...59`
## * `` -> `...60`
## * `` -> `...61`
## * `` -> `...62`
## * `` -> `...63`
## * `` -> `...64`
## * `` -> `...65`
## * `` -> `...66`
## * `` -> `...67`
## * `` -> `...68`
## * `` -> `...69`
## * `` -> `...70`
```

```
## * `` -> `...71`
## * `` -> `...72`
## * `` -> `...73`
## * `` -> `...74`
## * `` -> `...75`
## * `` -> `...76`
## * `` -> `...77`
## * `` -> `...78`
## * `` -> `...79`
## * `` -> `...80`
## * `` -> `...81`
## * `` -> `...82`
## * `` -> `...83`
## * `` -> `...84`
## * `` -> `...85`
## * `` -> `...86`
## * `` -> `...87`
## * `` -> `...88`
## * `` -> `...89`
## * `` -> `...90`
## * `` -> `...91`
## * `` -> `...92`
## * `` -> `...93`
## * `` -> `...94`
## * `` -> `...95`
## * `` -> `...96`
## * `` -> `...97`
## * `` -> `...98`
## * `` -> `...99`
## * `` -> `...100`
## * `` -> `...101`
## * `` -> `...102`
## * `` -> `...103`
## * `` -> `...104`
## * `` -> `...105`
## * `` -> `...106`
## * `` -> `...107`
## * `` -> `...108`
## * `` -> `...109`
## * `` -> `...110`
## * `` -> `...111`
## * `` -> `...112`
## * `` -> `...113`
## * `` -> `...114`
## * `` -> `...115`
## * `` -> `...116`
## * `` -> `...117`
## * `` -> `...118`
## * `` -> `...119`
## * `` -> `...120`
## * `` -> `...121`
## * `` -> `...122`
## * `` -> `...123`
## * `` -> `...124`
```

```
## * `` -> `...125`
## * `` -> `...126`
## * `` -> `...127`
## * `` -> `...128`
## * `` -> `...129`
## * `` -> `...130`
## * `` -> `...131`
## * `` -> `...132`
## * `` -> `...133`
## * `` -> `...134`
## * `` -> `...135`
## * `` -> `...136`
## * `` -> `...137`
## * `` -> `...138`
## * `` -> `...139`
## * `` -> `...140`
## * `` -> `...141`
## * `` -> `...142`
## * `` -> `...143`
## * `` -> `...144`
## * `` -> `...145`
## * `` -> `...146`
## * `` -> `...147`
## * `` -> `...148`
## * `` -> `...149`
## * `` -> `...150`
## * `` -> `...151`
## * `` -> `...152`
## * `` -> `...153`
## * `` -> `...154`
## * `` -> `...155`
## * `` -> `...156`
## * `` -> `...157`
## * `` -> `...158`
## * `` -> `...159`
## * `` -> `...160`
## * `` -> `...161`
## * `` -> `...162`
## * `` -> `...163`
## * `` -> `...164`
## * `` -> `...165`
## * `` -> `...166`
## * `` -> `...167`
## * `` -> `...168`
## * `` -> `...169`
## * `` -> `...170`
## * `` -> `...171`
## * `` -> `...172`
## * `` -> `...173`
## * `` -> `...174`
## * `` -> `...175`
## * `` -> `...176`
## * `` -> `...177`
## * `` -> `...178`
```

```
## * `` -> `...179`
## * `` -> `...180`
## * `` -> `...181`
## * `` -> `...182`
## * `` -> `...183`
## * `` -> `...184`
## * `` -> `...185`
## * `` -> `...186`
## * `` -> `...187`
## * `` -> `...188`
## * `` -> `...189`
## * `` -> `...190`
## * `` -> `...191`
## * `` -> `...192`
## * `` -> `...193`
## * `` -> `...194`
## * `` -> `...195`
## * `` -> `...196`
## * `` -> `...197`
## * `` -> `...198`
## * `` -> `...199`
## * `` -> `...200`
## * `` -> `...201`
## * `` -> `...202`
## * `` -> `...203`
## * `` -> `...204`
## * `` -> `...205`
## * `` -> `...206`
## * `` -> `...207`
## * `` -> `...208`
## * `` -> `...209`
## * `` -> `...210`
## * `` -> `...211`
## * `` -> `...212`
## * `` -> `...213`
## * `` -> `...214`
## * `` -> `...215`
## * `` -> `...216`
## * `` -> `...217`
## * `` -> `...218`
## * `` -> `...219`
## * `` -> `...220`
## * `` -> `...221`
## * `` -> `...222`
## * `` -> `...223`
## * `` -> `...224`
## * `` -> `...225`
## * `` -> `...226`
## * `` -> `...227`
## * `` -> `...228`
## * `` -> `...229`
## * `` -> `...230`
## * `` -> `...231`
## * `` -> `...232`
```

```
## * `` -> `...233`
## * `` -> `...234`
## * `` -> `...235`
## * `` -> `...236`
## * `` -> `...237`
## * `` -> `...238`
## * `` -> `...239`
## * `` -> `...240`
## * `` -> `...241`
## * `` -> `...242`
## * `` -> `...243`
## * `` -> `...244`
## * `` -> `...245`
## * `` -> `...246`
## * `` -> `...247`
## * `` -> `...248`
## * `` -> `...249`
## * `` -> `...250`
## * `` -> `...251`
## * `` -> `...252`
## * `` -> `...253`
## * `` -> `...254`
## * `` -> `...255`
## * `` -> `...256`
## * `` -> `...257`
## * `` -> `...258`
## * `` -> `...259`
## * `` -> `...260`
## * `` -> `...261`
## * `` -> `...262`
## * `` -> `...263`
## * `` -> `...264`
## * `` -> `...265`
## * `` -> `...266`
## * `` -> `...267`
## * `` -> `...268`
## * `` -> `...269`
## * `` -> `...270`
## * `` -> `...271`
## * `` -> `...272`
## * `` -> `...273`
## * `` -> `...274`
## * `` -> `...275`
## * `` -> `...276`
## * `` -> `...277`
## * `` -> `...278`
## * `` -> `...279`
## * `` -> `...280`
## * `` -> `...281`
## * `` -> `...282`
## * `` -> `...283`
## * `` -> `...284`
## * `` -> `...285`
## * `` -> `...286`
```

```
## * `` -> `...287`
## * `` -> `...288`
## * `` -> `...289`
## * `` -> `...290`
## * `` -> `...291`
## * `` -> `...292`
## * `` -> `...293`
## * `` -> `...294`
## * `` -> `...295`
## * `` -> `...296`
## * `` -> `...297`
## * `` -> `...298`
## * `` -> `...299`
## * `` -> `...300`
## * `` -> `...301`
## * `` -> `...302`
## * `` -> `...303`
## * `` -> `...304`
## * `` -> `...305`
## * `` -> `...306`
## * `` -> `...307`
## * `` -> `...308`
## * `` -> `...309`
## * `` -> `...310`
## * `` -> `...311`
## * `` -> `...312`
## * `` -> `...313`
## * `` -> `...314`
## * `` -> `...315`
## * `` -> `...316`
## * `` -> `...317`
## * `` -> `...318`
## * `` -> `...319`
## * `` -> `...320`
## * `` -> `...321`
## * `` -> `...322`
## * `` -> `...323`
## * `` -> `...324`
## * `` -> `...325`
## * `` -> `...326`
## * `` -> `...327`
## * `` -> `...328`
## * `` -> `...329`
## * `` -> `...330`
## * `` -> `...331`
## * `` -> `...332`
## * `` -> `...333`
## * `` -> `...334`
## * `` -> `...335`
## * `` -> `...336`
## * `` -> `...337`
## * `` -> `...338`
## * `` -> `...339`
## * `` -> `...340`
```

```
## * `` -> `...341`
## * `` -> `...342`
## * `` -> `...343`
## * `` -> `...344`
## * `` -> `...345`
## * `` -> `...346`
## * `` -> `...347`
## * `` -> `...348`
## * `` -> `...349`
## * `` -> `...350`
## * `` -> `...351`
## * `` -> `...352`
## * `` -> `...353`
## * `` -> `...354`
## * `` -> `...355`
## * `` -> `...356`
## * `` -> `...357`
## * `` -> `...358`
## * `` -> `...359`
## * `` -> `...360`
## * `` -> `...361`
## * `` -> `...362`
## * `` -> `...363`
## * `` -> `...364`
## * `` -> `...365`
## * `` -> `...366`
## * `` -> `...367`
## * `` -> `...368`
## * `` -> `...369`
## * `` -> `...370`
## * `` -> `...371`
## * `` -> `...372`
## * `` -> `...373`
## * `` -> `...374`
## * `` -> `...375`
## * `` -> `...376`
## * `` -> `...377`
## * `` -> `...378`
## * `` -> `...379`
## * `` -> `...380`
## * `` -> `...381`
## * `` -> `...382`
## * `` -> `...383`
## * `` -> `...384`
## * `` -> `...385`
## * `` -> `...386`
## * `` -> `...387`
## * `` -> `...388`
## * `` -> `...389`
## * `` -> `...390`
## * `` -> `...391`
## * `` -> `...392`
## * `` -> `...393`
## * `` -> `...394`
```

```
## * `` -> `...395`
## * `` -> `...396`
## * `` -> `...397`
## * `` -> `...398`
## * `` -> `...399`
## * `` -> `...400`
## * `` -> `...401`
## * `` -> `...402`
## * `` -> `...403`
## * `` -> `...404`
## * `` -> `...405`
## * `` -> `...406`
## * `` -> `...407`
## * `` -> `...408`
## * `` -> `...409`

## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 356772 Columns: 409
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (5): Id, StateAbbr, QuestionUno, PostText, CreatedUtc
## lgl (404): ...6, ...7, ...8, ...9, ...10, ...11, ...12, ...13, ...14, ...15,...
##
## 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.
```

```
questions <- read.csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest 2023
statesites <- read.csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest 20
subcategories <- read.csv(file = "/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest
```

```
## clean clients table
head(clients)
```

```
##   Id StateAbbr                            ClientUno     County   StateName
## 1  1        MS B731DA7A-A97A-4A2E-BF11-0A126A05AF70 Washington Mississippi
## 2  2        MS 8101DB7C-0CC8-4DE9-9E87-71137269AAD5   Harrison Mississippi
## 3  3        GA DD5F1EFC-8298-457F-BC09-CABBEB57A697     Barrow     Georgia
## 4  4        TX 21DC9E3D-D0AA-4B85-88BB-428F5BCF7F13    Lubbock       Texas
## 5  5        ID 3782D4F9-9B16-4BC8-ADA3-F85CEE200630      Boise       Idaho
## 6  6        CT F8C0C4C9-0555-4BAF-9324-B2A8DF4E7DF6  New Haven Connecticut
##   PostalCode     EthnicIdentity  Age Gender       MaritalStatus Veteran
## 1      73179                      NULL   NULL                NULL    NULL
## 2      39560          Caucasian NULL   Male Divorced or Widowed    NULL
## 3      30157 Latino or Hispanic   55   Male              Single    NULL
## 4      79416              Other   61 Female Married / remarried      No
## 5      83705          Caucasian   65   Male              Single      No
## 6       6511   African American NULL Female              Single    NULL
##   Imprisoned NumberInHousehold AnnualIncome AllowedIncome CheckingBalance
## 1         No                3        26000         37190             300
```

```
## 2         No                 2           52         25390              1
## 3         No                 5         3000         60790           NULL
## 4         No                 4        15600         48990            100
## 5         No                 1        15000         13590            580
## 6         No                 3        16800         37190             65
##   SavingsBalance InvestmentsBalance          CreatedUtc
## 1           NULL               NULL 2016-08-25 15:15:37
## 2              1               NULL 2016-08-26 00:55:17
## 3           NULL               NULL 2016-10-12 14:37:06
## 4           NULL               NULL 2017-01-01 19:57:07
## 5             25               NULL 2017-02-23 00:03:27
## 6           NULL               NULL 2016-08-29 18:54:49
```

```
# drop irrelevant information
clients <- subset(clients, select = -c(County, StateName,
                                       PostalCode))
# remove NULL rows
clients[clients == "NULL"] = NA
clients <- clients[rowSums(is.na(clients)) < 9, ]
# numeric type
clients$AnnualIncome <- as.numeric(clients$AnnualIncome)

##  clean questions table
head(questions)
```

```
##   Id StateAbbr                          QuestionUno
## 1  1        NC BB53B2A4-7C1B-45C3-9693-62047889317D
## 2  2        NC 0C49F841-2782-4313-961E-7882B96D0854
## 3  3        NC BFC28E08-B6B1-4185-A91A-89DF90816BE4
## 4  4        NC 45B24109-243A-489B-B0D2-2F36C2A369FE
## 5  5        NC C4707F71-FE4A-4154-8174-AF00BEE2B361
## 6  6        NC C157FEFD-9F47-4288-B853-E1E7918A568F
##                            CategoryUno                     Category
## 1 C932D197-7F43-4BCF-BE0B-FE5F5DBCB9EF          Family and Children
## 2 C8E1275A-6C42-42E4-9B7D-231BAC438FC9            Individual Rights
## 3 C8E1275A-6C42-42E4-9B7D-231BAC438FC9            Individual Rights
## 4 C932D197-7F43-4BCF-BE0B-FE5F5DBCB9EF          Family and Children
## 5 F6016FBD-9501-4448-BCD3-BBB0E834E33B Consumer Financial Questions
## 6 C932D197-7F43-4BCF-BE0B-FE5F5DBCB9EF          Family and Children
##                         SubcategoryUno
## 1 E9F87919-0CD5-4D9D-AAC6-3CBC49132F1D
## 2 73A59926-BA0B-44EC-ABF3-F4C9181EEA83
## 3 73A59926-BA0B-44EC-ABF3-F4C9181EEA83
## 4 E9F87919-0CD5-4D9D-AAC6-3CBC49132F1D
## 5 73FE486E-2B51-456A-A6D9-78A7D05802AA
## 6 E9F87919-0CD5-4D9D-AAC6-3CBC49132F1D
##                                     Subcategory
## 1 Family/Divorce/Custody/Visitation/Child Support
## 2                     Civil/Constitutional Rights
## 3                     Civil/Constitutional Rights
## 4 Family/Divorce/Custody/Visitation/Child Support
## 5                   Bankruptcy, Debts & Purchases
## 6 Family/Divorce/Custody/Visitation/Child Support
##                         AskedByClientUno          AskedOnUtc
```

```
## 1 9C35423C-914A-42BC-8EBC-A370E76E2178 2019-06-19 22:10:06
## 2 DF636AEE-7D08-4F90-9A5C-0A2A726C6B59 2019-06-19 23:34:36
## 3 DF636AEE-7D08-4F90-9A5C-0A2A726C6B59 2019-06-19 23:57:18
## 4 C9C3FA40-0D60-49B9-97F1-63AFEFAFA484 2019-06-20 02:33:14
## 5 ED7DEA10-CDB7-4A05-891D-889D1EA710BE 2019-06-20 17:04:47
## 6 19A1157F-BAF1-437B-84FA-41C5FDF6DE97 2019-06-20 18:11:56
##                     TakenByAttorneyUno          TakenOnUtc
## 1                                 NULL                NULL
## 2                                 NULL                NULL
## 3 FA34142B-1575-4720-981C-2D28C3560137 2019-06-22 02:37:16
## 4 FA34142B-1575-4720-981C-2D28C3560137 2019-07-27 18:04:08
## 5                                 NULL                NULL
## 6                                 NULL                NULL
##                   ClosedByAttorneyUno         ClosedOnUtc       LegalDeadline
## 1                                 NULL 2019-08-13 15:09:06 2019-07-08 00:00:00
## 2                                 NULL 2019-08-13 15:09:14                NULL
## 3 FA34142B-1575-4720-981C-2D28C3560137 2019-06-24 19:02:35                NULL
## 4 FA34142B-1575-4720-981C-2D28C3560137 2019-07-31 05:49:18                NULL
## 5                                 NULL 2019-08-13 15:09:20                NULL
## 6                                 NULL 2019-08-13 15:10:26                NULL
```

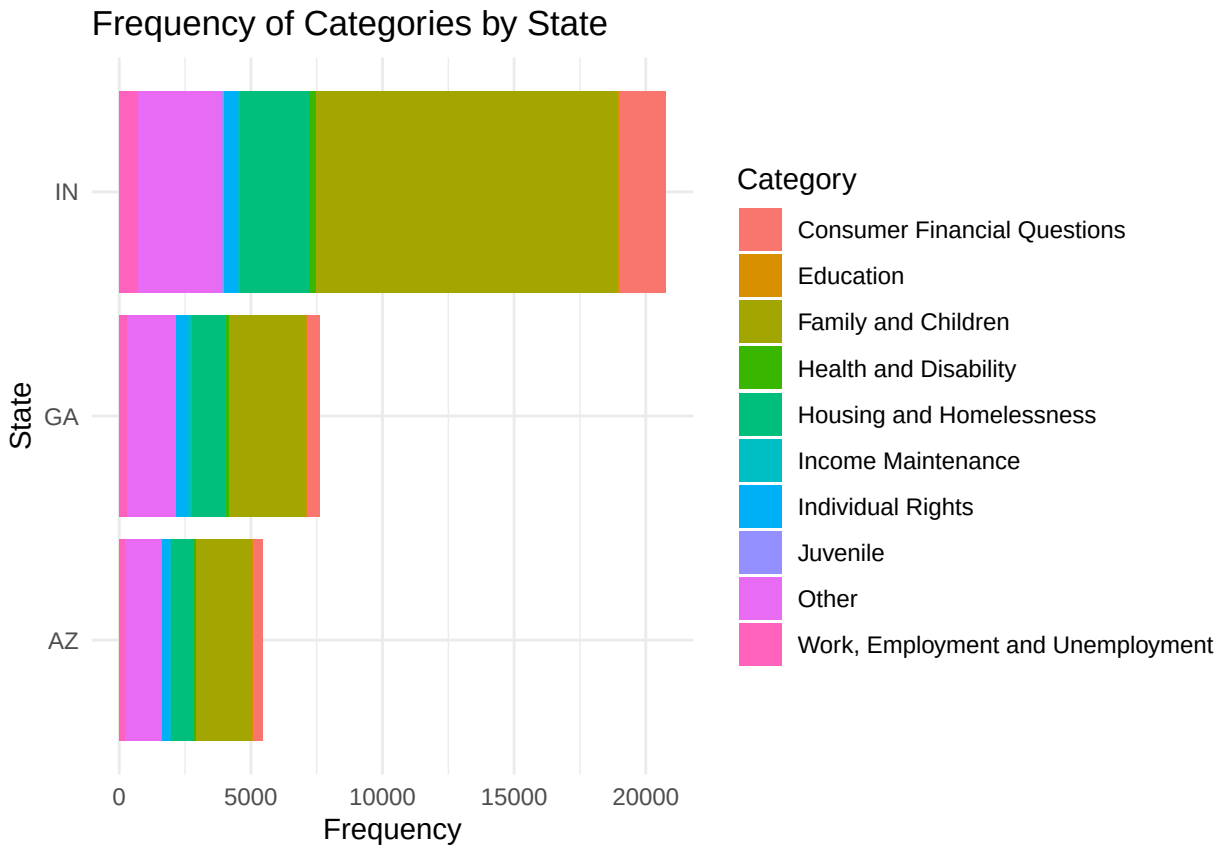
```
questions[questions == "NULL"] = NA

### VISUALIZATION

## Frequency of categories for each state
# Plot 1: Stacked Bar Graph
# group the data by state and category, count the number of occurrences
questions_subset <- subset(questions, StateAbbr %in% c("IN", "AZ", "GA"))
count <- questions_subset %>% group_by(StateAbbr, Category) %>% summarise(count = n())
```

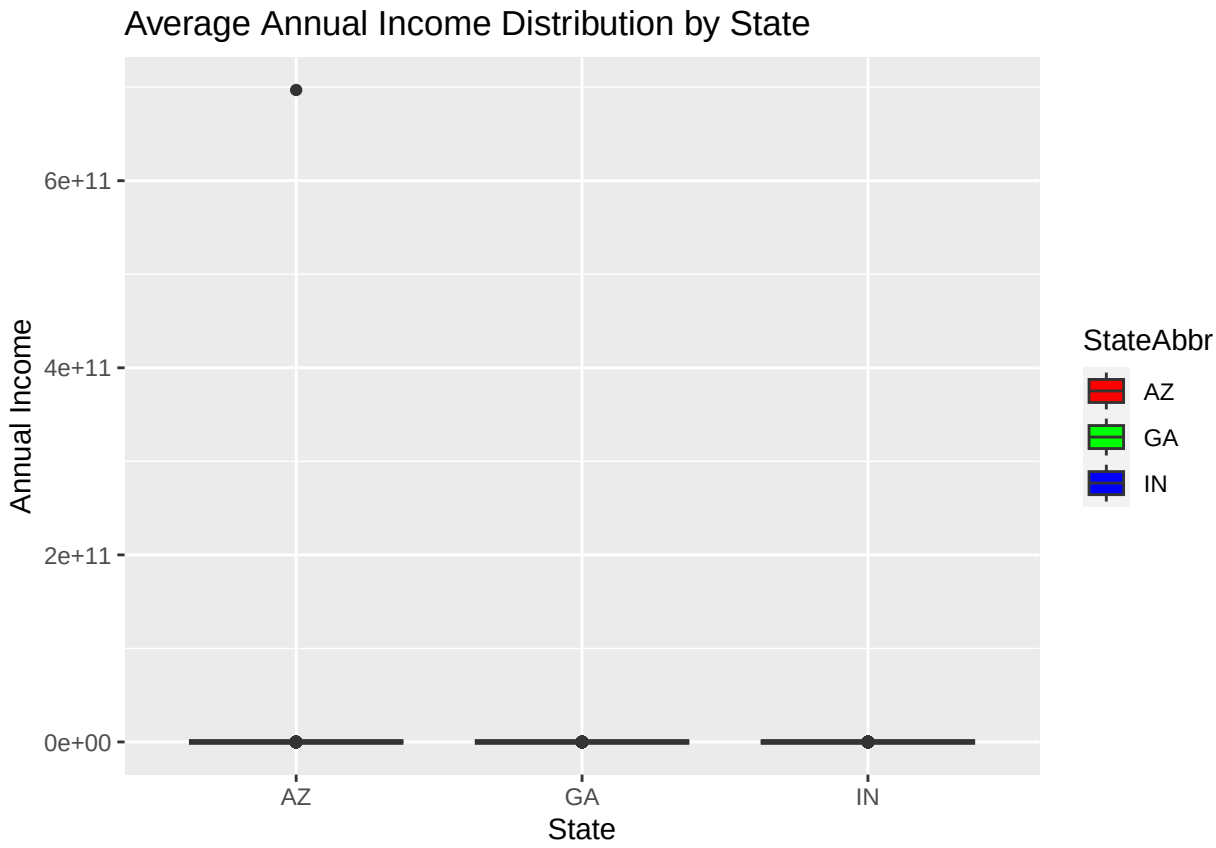
```
## `summarise()` has grouped output by 'StateAbbr'. You can override using the
## `.groups` argument.
```

```
# plot
ggplot(count, aes(x = count, y = StateAbbr, fill = Category)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(x = "Frequency", y = "State", fill = "Category",
       title = "Frequency of Categories by State")
```




```
## Income Data for each State
clients_subset <- subset(clients, StateAbbr %in% c("IN", "AZ", "GA"))
#clients_subset <- clients_subset[clients_subset$AnnualIncome < 200000,]
ggplot(clients_subset, aes(x=StateAbbr, y=as.numeric(AnnualIncome), fill=StateAbbr)) +
  geom_boxplot() +
  scale_fill_manual(values = rainbow(length(unique(clients_subset$StateAbbr)))) +
  xlab("State") + ylab("Annual Income") +
  ggtitle("Average Annual Income Distribution by State")
```

```
## Warning: Removed 3113 rows containing non-finite values (‘stat_boxplot()‘).
```




```
options(scipen=999)
```