Project_1

Zachary D'Urso

2024-11-18

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 (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:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Question Restated:

How can our company attract and retain top data science talent while battling the recession despite a highly competitive job market? What salary range should we offer for a full-time data scientist in the U.S. compared to offshore options?

Alternative Questions:

What is a competitive salary range for data scientists at various experience levels?

How do salaries for data scientists vary based on a company's size, location, and remote work ratio?

What are the cost differences between hiring a U.S.-based data scientist versus an offshore data scientist with the same experience level?

How might remote work impact data scientist salaries across offshore options?

```
#file.choose() # find location of file

library(tidyverse) # load library
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```r
data <- read.csv("C:\\Users\\zdurs\\Downloads\\r project data.csv") # Load/read dataset
```

```r
data <- data %>% #create dataset
  mutate(
    experience_level = factor(experience_level, levels = c("EN", "MI", "SE", "EX")), #change var
iable to factor
    employment_type = factor(employment_type, levels = c("PT", "FT", "CT", "FL")), #change varia
ble to factor
    remote_ratio = factor(remote_ratio, levels = c(0, 50, 100), labels = c("No Remote", "Partial
ly Remote", "Fully Remote")), #change variable to factor
    company_size = factor(company_size, levels = c("S", "M", "L")) #change variable to factor
  )
```

```r
colSums(is.na(data)) # check for NAs within each variable in the dataset
```

```
##                  X          work_year   experience_level    employment_type 
##                  0                  0                  0                  0 
##          job_title             salary    salary_currency      salary_in_usd 
##                  0                  0                  0                  0 
## employee_residence       remote_ratio   company_location       company_size 
##                  0                  0                  0                  0
```

```r
# Summary for the entire dataset
summary(data)
```

```
##        X           work_year    experience_level employment_type
##  Min.   :  0.0   Min.   :2020   EN: 88           PT: 10         
##  1st Qu.:151.5   1st Qu.:2021   MI:213           FT:588         
##  Median :303.0   Median :2022   SE:280           CT:  5         
##  Mean   :303.0   Mean   :2021   EX: 26           FL:  4         
##  3rd Qu.:454.5   3rd Qu.:2022                                   
##  Max.   :606.0   Max.   :2022                                   
##   job_title             salary         salary_currency    salary_in_usd   
##  Length:607         Min.   :    4000   Length:607         Min.   :  2859  
##  Class :character   1st Qu.:   70000   Class :character   1st Qu.: 62726  
##  Mode  :character   Median :  115000   Mode  :character   Median :101570  
##                     Mean   :  324000                      Mean   :112298  
##                     3rd Qu.:  165000                      3rd Qu.:150000  
##                     Max.   :30400000                      Max.   :600000  
##  employee_residence           remote_ratio company_location   company_size
##  Length:607         No Remote       :127   Length:607         S: 83       
##  Class :character   Partially Remote: 99   Class :character   M:326       
##  Mode  :character   Fully Remote    :381   Mode  :character   L:198       
##                                                                           
##                                                                           
## 
```

```
# To examine each variable individually we can summarize like this
summary(data$work_year)          # Years in the dataset; Years range from 2020-2022
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2020    2021    2022    2021    2022    2022
```

```
summary(data$experience_level)  # Counts for each experience level in the dataset; 4 types of ex
perience level, SE has the highest count
```

```
##  EN  MI  SE  EX
##  88 213 280  26
```

```
summary(data$employment_type)   # Counts for employment types; 4 types of employment type, FT ha
s the highest count
```

```
##  PT  FT  CT  FL
##  10 588   5   4
```

```
summary(data$salary)            # Range/distribution of salary; ranges from 4,000 to 30,400,000
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##     4000    70000   115000   324000   165000 30400000
```

```
summary(data$salaryinusd)       # Range/distribution of salaries in USD; ranges from 2,859 to 60
0,000
```

```
## Length  Class   Mode
##      0   NULL   NULL
```

```
summary(data$employee_residence) # Distribution of employee residence
```

```
##    Length     Class      Mode
##       607 character character
```

```
summary(data$remote_ratio)      # Counts for remote work; 3 types of remote work, fully remote h
as the highest count
```

```
##        No Remote Partially Remote     Fully Remote
##              127               99              381
```

```
summary(data$company_location)  # Distribution of company location
```

```
##    Length     Class      Mode 
##       607 character character
```

```
summary(data$company_size)      # Company size; 3 types of company size; M has the highest count
```

```
##   S   M   L 
##  83 326 198
```

```
# Count unique values for each variable (which type and how many of them for each variable)
sapply(data, function(x) {
  list(Variable_Type = class(x), Unique_Values = length(unique(x)))
})
```

```
##               X         work_year experience_level employment_type job_title  
## Variable_Type "integer" "integer" "factor"         "factor"        "character"
## Unique_Values 607       3         4                4               50         
##               salary    salary_currency salary_in_usd employee_residence
## Variable_Type "integer" "character"     "integer"     "character"       
## Unique_Values 272       17              369           57                
##               remote_ratio company_location company_size
## Variable_Type "factor"     "character"      "factor"    
## Unique_Values 3            50               3
```

```
# Example tables: Frequency distribution for each variable
table(data$company_size)
```

```
## 
##   S   M   L 
##  83 326 198
```

```
table(data$remote_ratio)
```

```
## 
##        No Remote Partially Remote     Fully Remote 
##              127               99              381
```

```
table(data$experience_level)
```

```
## 
##  EN  MI  SE  EX 
##  88 213 280  26
```

```
table(data$work_year)
```

```
##
## 2020 2021 2022
##   72  217  318
```

```
table(data$employment_type)
```

```
##
##  PT  FT  CT  FL
##  10 588   5   4
```

```
table(data$job_title)
```

```
##
##            3D Computer Vision Researcher
##                                        1
##                             AI Scientist
##                                        7
##                       Analytics Engineer
##                                        4
##                   Applied Data Scientist
##                                        5
##        Applied Machine Learning Scientist
##                                        4
##                          BI Data Analyst
##                                        6
##                       Big Data Architect
##                                        1
##                        Big Data Engineer
##                                        8
##                    Business Data Analyst
##                                        5
##                      Cloud Data Engineer
##                                        2
##                 Computer Vision Engineer
##                                        6
##        Computer Vision Software Engineer
##                                        3
##                             Data Analyst
##                                       97
##                  Data Analytics Engineer
##                                        4
##                      Data Analytics Lead
##                                        1
##                   Data Analytics Manager
##                                        7
##                           Data Architect
##                                       11
##                            Data Engineer
##                                      132
##                 Data Engineering Manager
##                                        5
##                  Data Science Consultant
##                                        7
##                    Data Science Engineer
##                                        3
##                     Data Science Manager
##                                       12
##                           Data Scientist
##                                      143
##                          Data Specialist
##                                        1
##             Director of Data Engineering
##                                        2
##                 Director of Data Science
```

```
##                                           7
##                                ETL Developer
##                                           2
##                        Finance Data Analyst
##                                           1
##                      Financial Data Analyst
##                                           2
##                                Head of Data
##                                           5
##                        Head of Data Science
##                                           4
##                    Head of Machine Learning
##                                           1
##                           Lead Data Analyst
##                                           3
##                          Lead Data Engineer
##                                           6
##                         Lead Data Scientist
##                                           3
##            Lead Machine Learning Engineer
##                                           1
##                Machine Learning Developer
##                                           3
##                 Machine Learning Engineer
##                                          41
## Machine Learning Infrastructure Engineer
##                                           3
##                  Machine Learning Manager
##                                           1
##                Machine Learning Scientist
##                                           8
##                    Marketing Data Analyst
##                                           1
##                               ML Engineer
##                                           6
##                              NLP Engineer
##                                           1
##                    Principal Data Analyst
##                                           2
##                   Principal Data Engineer
##                                           3
##                  Principal Data Scientist
##                                           7
##                      Product Data Analyst
##                                           2
##                        Research Scientist
##                                          16
##                      Staff Data Scientist
##                                           1
```

```
table(data$salary)
```

```
##
##    4000    8000    8760    9272   10000   12000   13400   14000
##       2       1       1       1       2       3       1       1
##   18000   19000   20000   21000   21600   21844   22000   24000
##       1       1       6       2       1       1       1       2
##   25000   28500   29000   30000   31000   32000   34000   35000
##       1       1       1       5       1       1       2       4
##   37000   37456   38400   39600   40000   40900   41000   42000
##       1       1       2       1       8       1       1       3
##   43200   44000   45000   45760   48000   50000   51400   51999
##       2       1       6       1       4      13       1       1
##   52000   52500   52800   53000   54000   55000   56000   57000
##       2       1       1       2       3       6       1       1
##   58000   59000   60000   61300   61500   62000   63900   65000
##       4       3      15       2       1       2       1       7
##   65720   66500   67000   68000   69000   69600   69999   70000
##       1       1       2       1       1       1       1      10
##   70500   72000   72500   73000   74000   75000   76760   78000
##       1       1       1       1       1      11       2       1
##   80000   81000   81666   82500   82900   84900   85000   87000
##      18       2       2       2       1       1       7       2
##   88000   90000   90320   90700   91000   93000   93150   93700
##       2      10       5       1       1       1       1       2
##   95000   95550   98000   99000   99050   99100   99360  100000
##       2       1       1       2       1       1       1      18
##  100800  101570  102000  102100  103000  104890  105000  105400
##       1       1       1       2       1       1       5       1
##  106000  106260  108000  108800  109000  109280  110000  110500
##       1       2       1       1       1       2       8       1
##  110925  111775  112000  112300  112900  113000  115000  115500
##       1       1       1       1       4       1       6       1
##  115934  116000  116150  118000  120000  120160  120500  120600
##       2       1       1       1      15       1       1       1
##  121000  123000  124190  125000  126000  126500  128875  129000
##       1       3       1       4       1       2       2       1
##  130000  130800  132000  132320  135000  136000  136600  136620
##      11       1       1       3       9       1       1       1
##  136994  137141  138000  138350  138600  140000  140250  140400
##       1       1       1       1       1       9       1       3
##  141300  144000  144854  145000  146000  147000  147800  148000
##       1       3       1       2       1       2       1       1
##  150000  150075  150260  151000  152000  152500  153000  154000
##      14       1       1       1       1       1       2       1
##  154600  155000  156600  157000  158200  159000  159500  160000
##       2       3       1       2       1       1       1       9
##  160080  161342  164000  164996  165000  165220  165400  167000
##       2       1       1       2       4       1       2       2
##  167875  168000  170000  174000  175000  175100  176000  177000
##       1       1       8       2       2       1       1       1
##  180000  181940  183600  184700  185000  185100  188000  189650
##       8       1       1       1       2       1       1       2
##  190000  190200  192400  192564  192600  195000  200000  200100
```

```
##        1        1        1        1        1        1       10        1
##   205300   206699   208775   209100   210000   211500   213120   214000
##        3        1        1        2        5        1        1        1
##   215300   216000   220000   220110   224000   225000   230000   235000
##        2        1        3        2        1        2        3        3
##   240000   241000   242000   243900   250000   256000   260000   266400
##        2        1        1        1        4        1        2        1
##   270000   276000   299000   300000   324000   325000   380000   400000
##        1        1        1        1        1        1        1        1
##   405000   412000   416000   420000   423000   435000   450000   600000
##        1        1        1        1        2        1        4        1
##   700000   720000  1200000  1250000  1335000  1400000  1450000  1600000
##        1        1        1        1        1        2        1        1
##  1672000  1799997  2100000  2200000  2250000  2400000  2500000  3000000
##        1        1        1        1        1        1        1        2
##  4000000  4450000  4900000  6000000  7000000  8500000 11000000 30400000
##        1        1        1        1        2        1        2        1
```

```
table(data$salary_in_usd)
```

```
##
##   2859   4000   5409   5679   5707   5882   6072   8000   9272   9466  10000
##      1      2      1      1      1      1      2      1      1      1      2
##  10354  12000  12103  12901  13400  15966  16228  16904  18000  18053  18442
##      1      3      1      1      1      1      1      1      1      1      2
##  18907  19609  20000  20171  21637  21669  21844  21983  22611  24000  24342
##      1      1      5      1      1      1      1      1      1      1      1
##  24823  25000  25532  26005  28016  28369  28399  28476  28609  29751  30428
##      2      1      1      1      1      1      1      1      1      1      1
##  31615  31875  32974  33511  33808  35590  35735  36259  36643  37236  37300
##      1      1      3      1      1      1      1      1      1      1      1
##  37825  38400  38776  39263  39916  40000  40038  40189  40481  40570  41689
##      1      1      1      2      1      1      1      1      1      1      1
##  42000  42197  43331  43966  45391  45618  45760  45807  45896  46597  46759
##      1      1      1      2      1      1      1      3      1      1      1
##  46809  47282  47899  48000  49268  49461  49646  50000  50180  51064  51321
##      1      1      1      1      1      2      1      5      1      1      1
##  51519  52000  52351  52396  53192  54000  54094  54238  54742  54957  55000
##      1      1      3      1      1      1      1      1      1      3      2
##  56000  56256  56738  58000  58035  58255  58894  59102  59303  60000  60757
##      1      1      1      3      1      1      1      2      1      5      1
##  61300  61467  61896  62000  62649  62651  62726  63711  63810  63831  63900
##      2      1      1      1      1      1      2      1      1      2      1
##  64849  65000  65013  65438  65949  66022  66265  67000  68147  68428  69000
##      1      2      1      3      2      1      1      1      1      1      1
##  69336  69741  69999  70000  70139  70500  70912  71444  71786  71982  72000
##      1      2      1      2      1      1      1      1      1      1      1
##  72212  72500  73000  74000  74130  75000  75774  76833  76940  76958  77364
##      1      1      1      1      1      4      1      3      2      1      1
##  77684  78000  78526  78791  79039  79197  79833  80000  81000  81666  82500
##      1      1      4      2      1      1      2      8      1      2      1
##  82528  82744  82900  84900  85000  86703  87000  87425  87738  87932  88654
##      2      1      1      1      4      1      1      1      1      4      3
##  89294  90000  90320  90700  90734  91000  91237  91614  93000  93150  93427
##      1      6      5      1      2      1      1      2      1      1      1
##  93700  94564  94665  95550  95746  96113  96282  98000  98158  99000  99050
##      2      1      1      1      1      1      1      1      3      2      1
##  99100  99360  99703 100000 100800 101570 102100 102839 103000 103160 103691
##      1      1      1     15      1      1      2      1      1      1      1
## 104702 104890 105000 105400 106000 106260 108800 109000 109024 109280 110000
##      2      1      5      1      1      2      1      1      1      2      5
## 110037 110500 110925 111775 112000 112300 112872 112900 113000 113476 114047
##      1      1      1      1      1      1      1      4      1      1      1
## 115000 115500 115934 116000 116150 116914 117104 117789 118000 118187 119059
##      5      1      2      1      1      1      1      2      1      1      1
## 120000 120160 120600 122346 123000 124190 124333 125000 126000 126500 127221
##     12      1      1      1      3      1      1      3      1      2      1
## 128875 129000 130000 130026 130800 132000 132320 135000 136000 136600 136620
##      2      1      8      1      1      1      3      9      1      1      1
## 136994 137141 138000 138350 138600 140000 140250 140400 141300 141846 144000
##      1      1      1      1      1      8      1      3      1      1      3
## 144854 145000 146000 147000 147800 148261 150000 150075 150260 151000 152000
```

```
##      1      2      1      1      1      1     12      1      1      1      1
## 152500 153000 153667 154000 154600 155000 156600 157000 158200 159000 160000
##      1      2      1      1      2      3      1      1      1      1      8
## 160080 161342 162674 164000 164996 165000 165220 165400 167000 167875 168000
##      2      1      1      1      2      4      1      2      2      1      1
## 170000 173762 174000 175000 175100 176000 177000 180000 181940 183228 183600
##      8      1      2      2      1      1      1      5      1      1      1
## 184700 185000 185100 187442 188000 189650 190000 190200 192400 192564 192600
##      1      2      1      1      1      2      1      1      1      1      1
## 195000 196979 200000 200100 205300 206699 208775 209100 210000 211500 213120
##      1      1     10      1      3      1      1      2      5      1      1
## 214000 215300 216000 220000 220110 224000 225000 230000 235000 240000 241000
##      1      2      1      3      2      1      2      3      2      1      1
## 242000 243900 250000 256000 260000 266400 270000 276000 324000 325000 380000
##      1      1      2      1      2      1      1      1      1      1      1
## 405000 412000 416000 423000 450000 600000
##      1      1      1      1      2      1
```

```
table(data$salary_currency)
```

```
##
## AUD BRL CAD CHF CLP CNY DKK EUR GBP HUF INR JPY MXN PLN SGD TRY USD
##   2   2  18   1   1   2   2  95  44   2  27   3   2   3   2   3 398
```

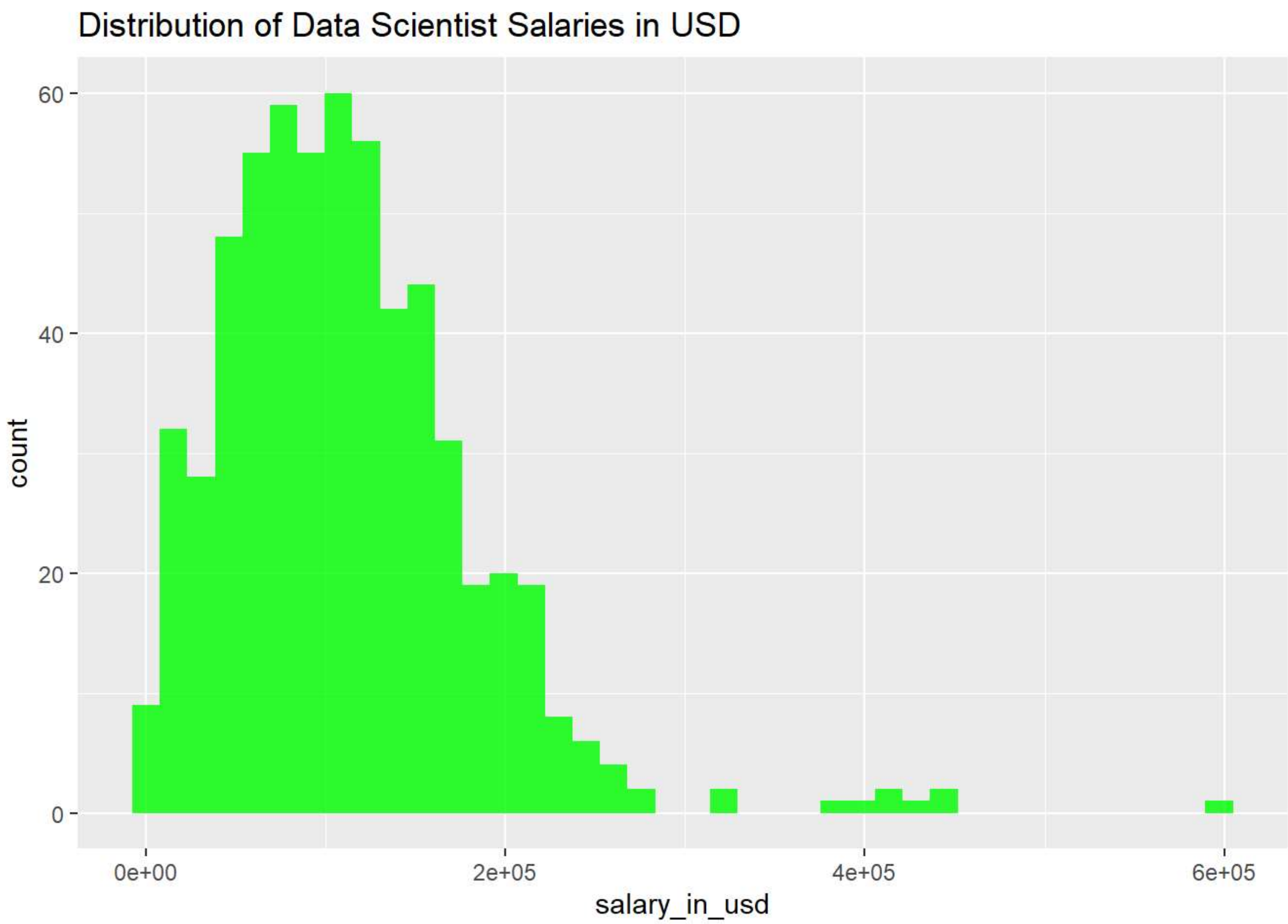
```
table(data$employee_residence)
```

```
##
##  AE  AR  AT  AU  BE  BG  BO  BR  CA  CH  CL  CN  CO  CZ  DE  DK  DZ  EE  ES  FR
##   3   1   3   3   2   1   1   6  29   1   1   1   1   1  25   2   1   1  15  18
##  GB  GR  HK  HN  HR  HU  IE  IN  IQ  IR  IT  JE  JP  KE  LU  MD  MT  MX  MY  NG
##  44  13   1   1   1   2   1  30   1   1   4   1   7   1   1   1   1   2   1   2
##  NL  NZ  PH  PK  PL  PR  PT  RO  RS  RU  SG  SI  TN  TR  UA  US  VN
##   5   1   1   6   4   1   6   2   1   4   2   2   1   3   1 332   3
```

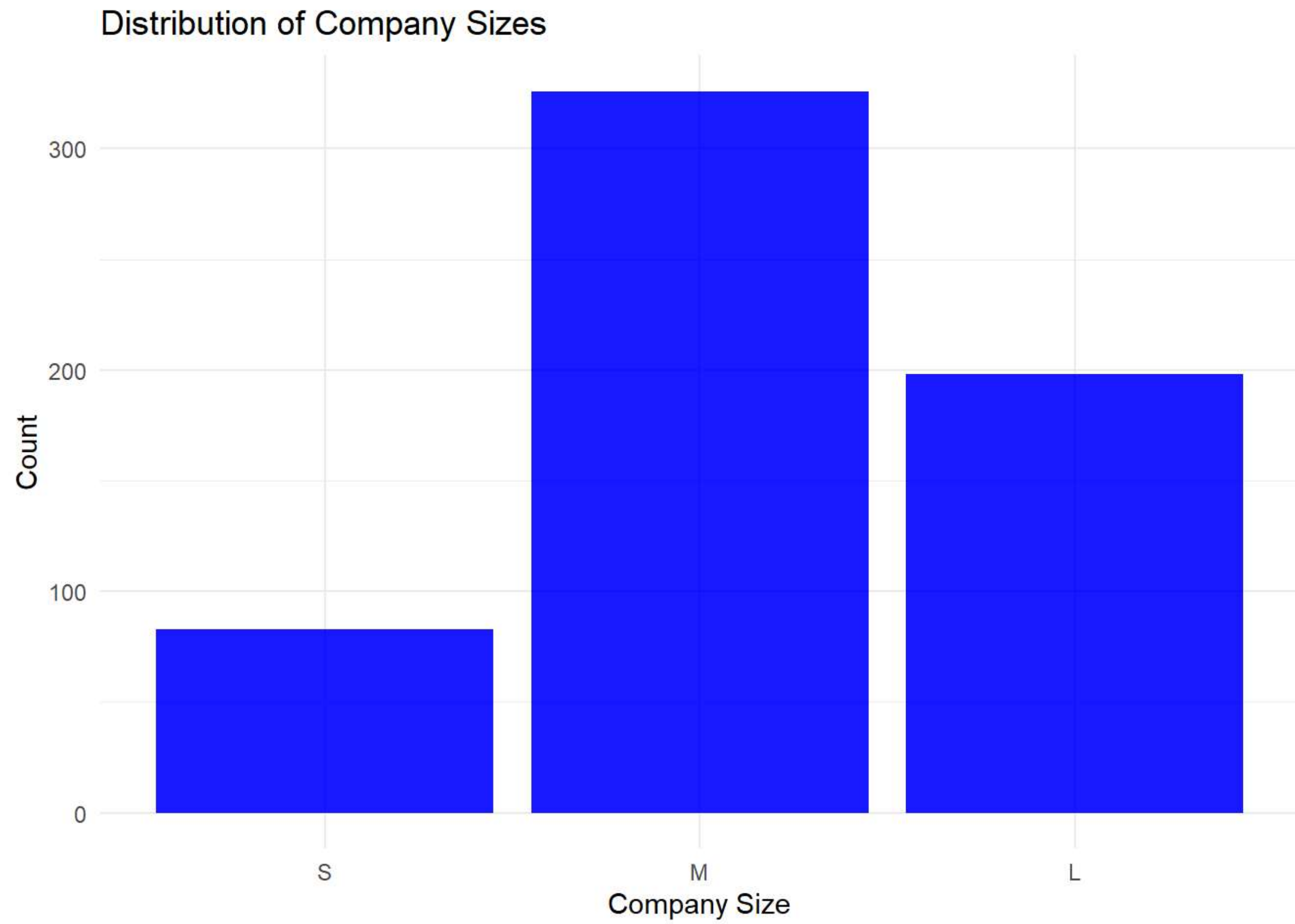
```
table(data$company_location)
```

```
##
##  AE  AS  AT  AU  BE  BR  CA  CH  CL  CN  CO  CZ  DE  DK  DZ  EE  ES  FR  GB  GR
##   3   1   4   3   2   3  30   2   1   2   1   2  28   3   1   1  14  15  47  11
##  HN  HR  HU  IE  IL  IN  IQ  IR  IT  JP  KE  LU  MD  MT  MX  MY  NG  NL  NZ  PK
##   1   1   1   1   1  24   1   1   2   6   1   3   1   1   3   1   2   4   1   3
##  PL  PT  RO  RU  SG  SI  TR  UA  US  VN
##   4   4   1   2   1   2   3   1 355   1
```

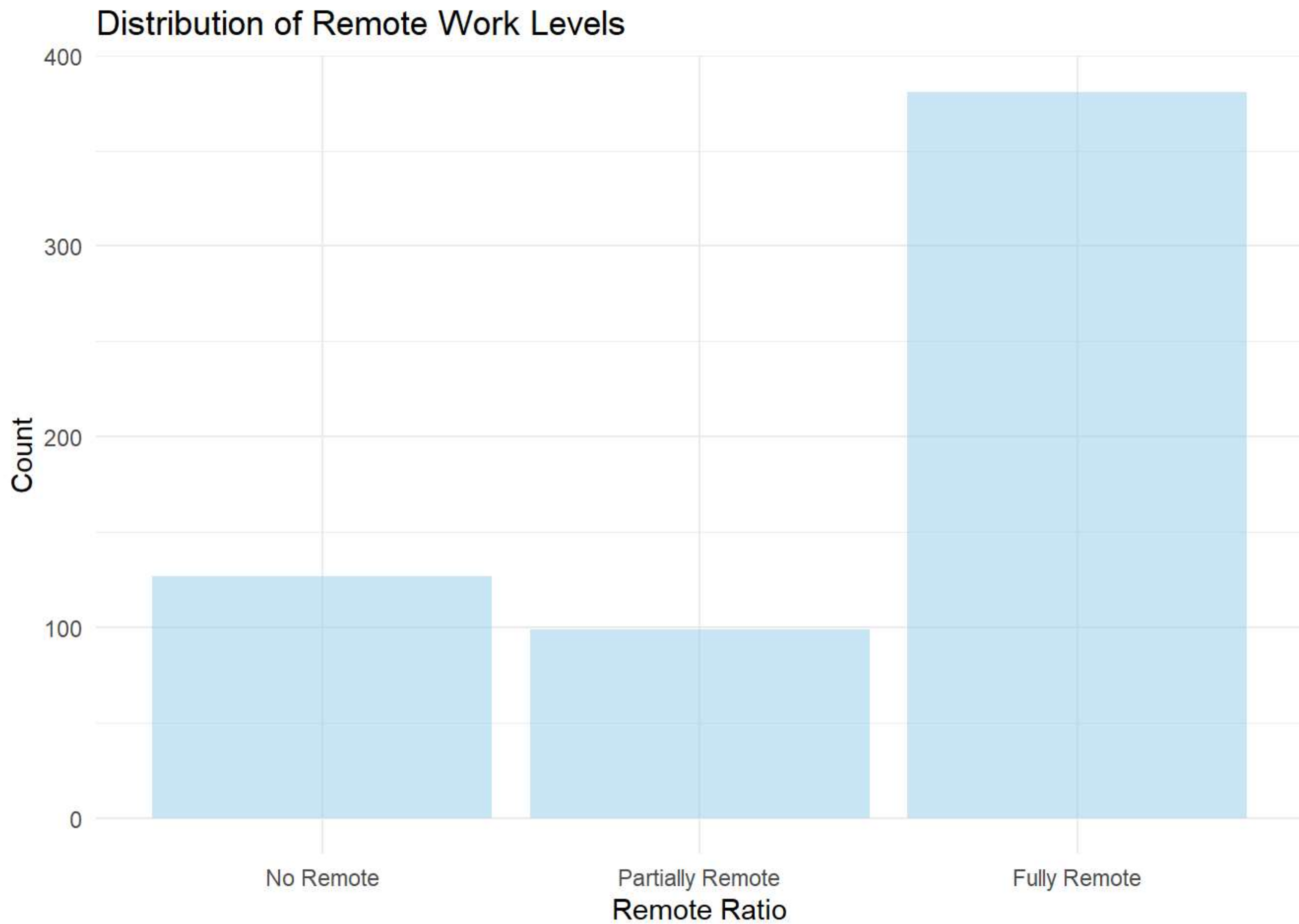
```r
# Example plot
library(ggplot2) #load libraries
library(dplyr)
ggplot(data, aes(x = salary_in_usd)) + #plot salary in USD
  geom_histogram(bins = 40, fill = "green", alpha = 0.8) + #create a histogram
  labs(title = "Distribution of Data Scientist Salaries in USD") #label the title
```



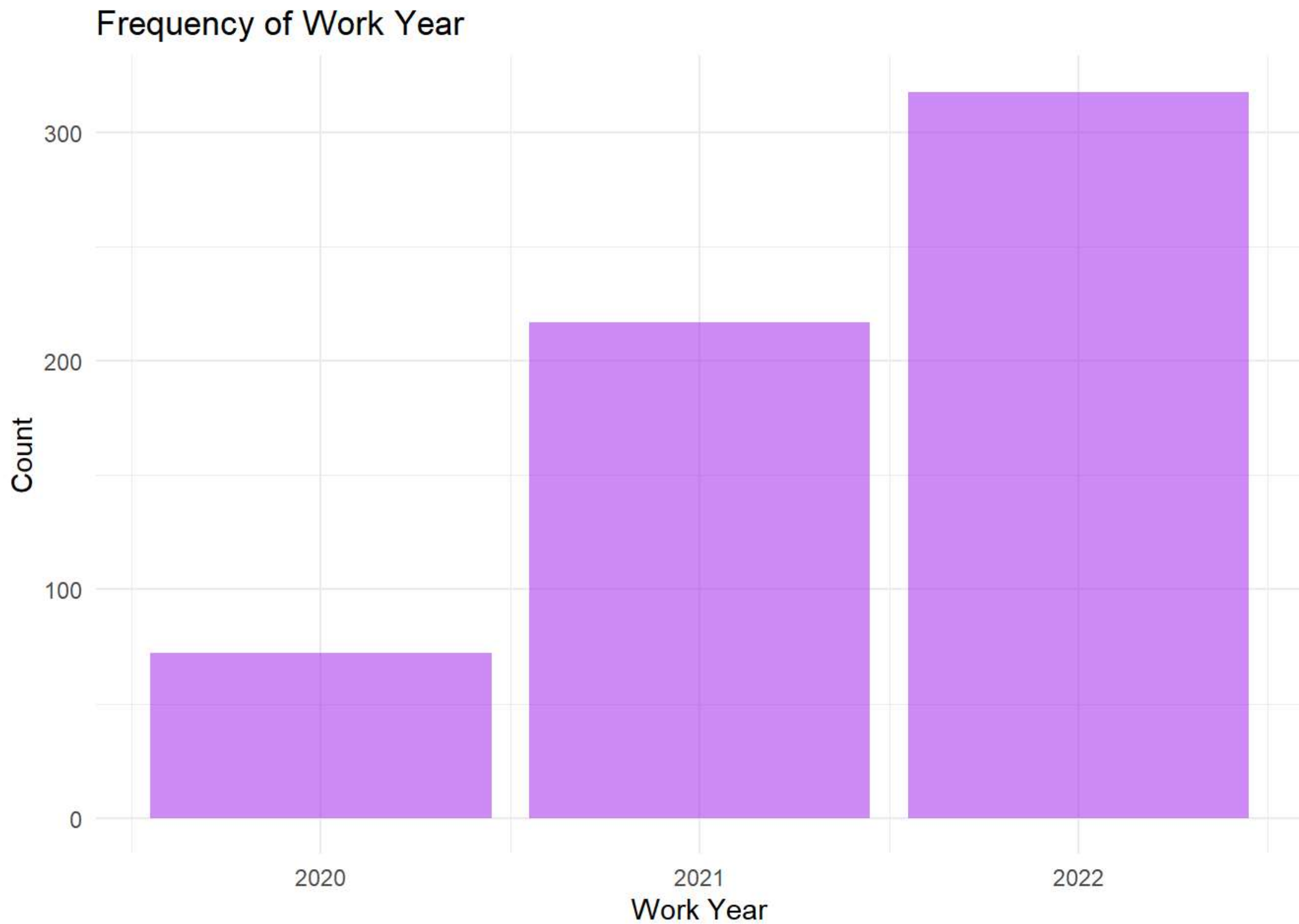

```r
ggplot(data, aes(x = company_size)) +
  geom_bar(fill = "blue", alpha = 0.9) +
  labs(title = "Distribution of Company Sizes", x = "Company Size", y = "Count") +
  theme_minimal()
```



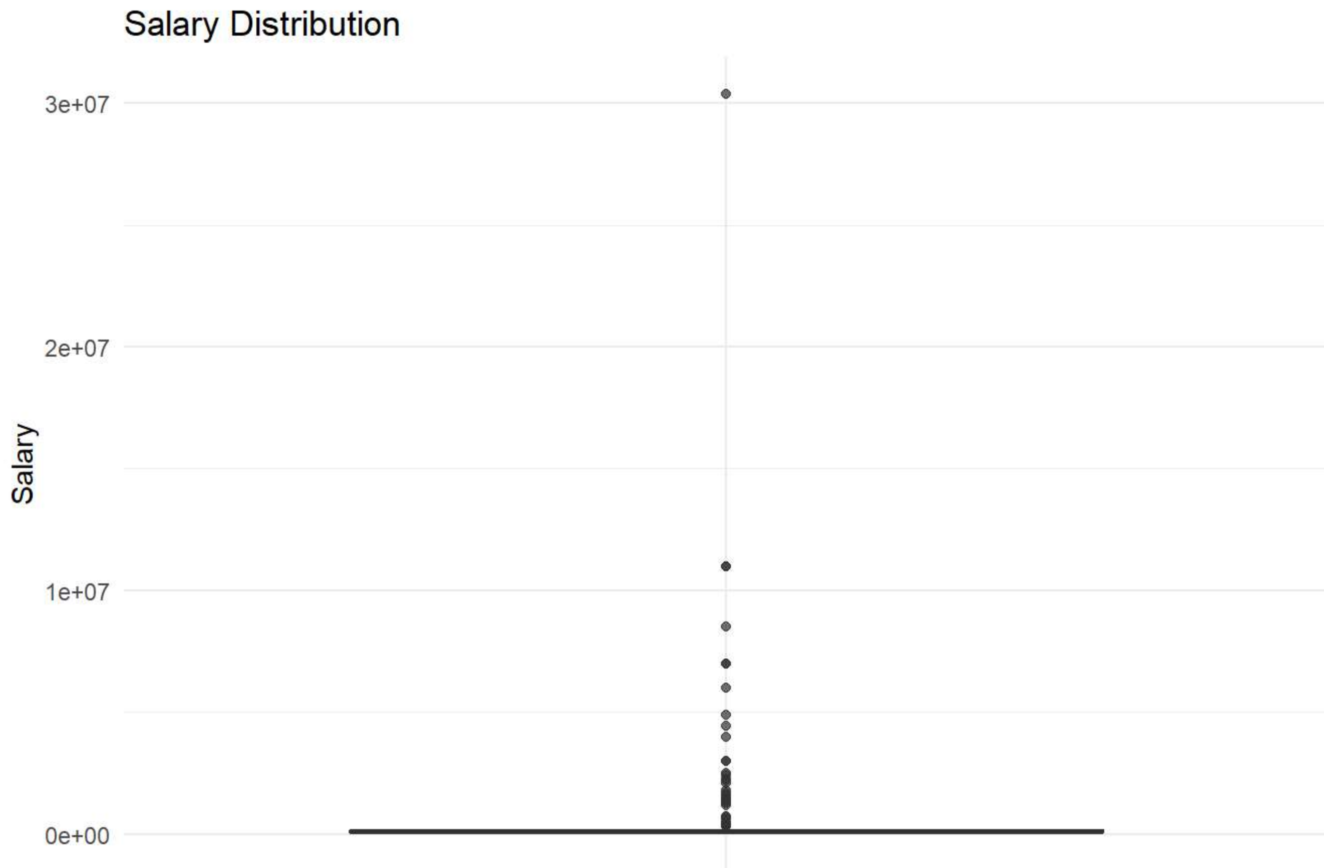

```
ggplot(data, aes(x = factor(remote_ratio))) +
  geom_bar(fill = "skyblue", alpha = 0.45) +
  labs(title = "Distribution of Remote Work Levels", x = "Remote Ratio", y = "Count") +
  theme_minimal()
```



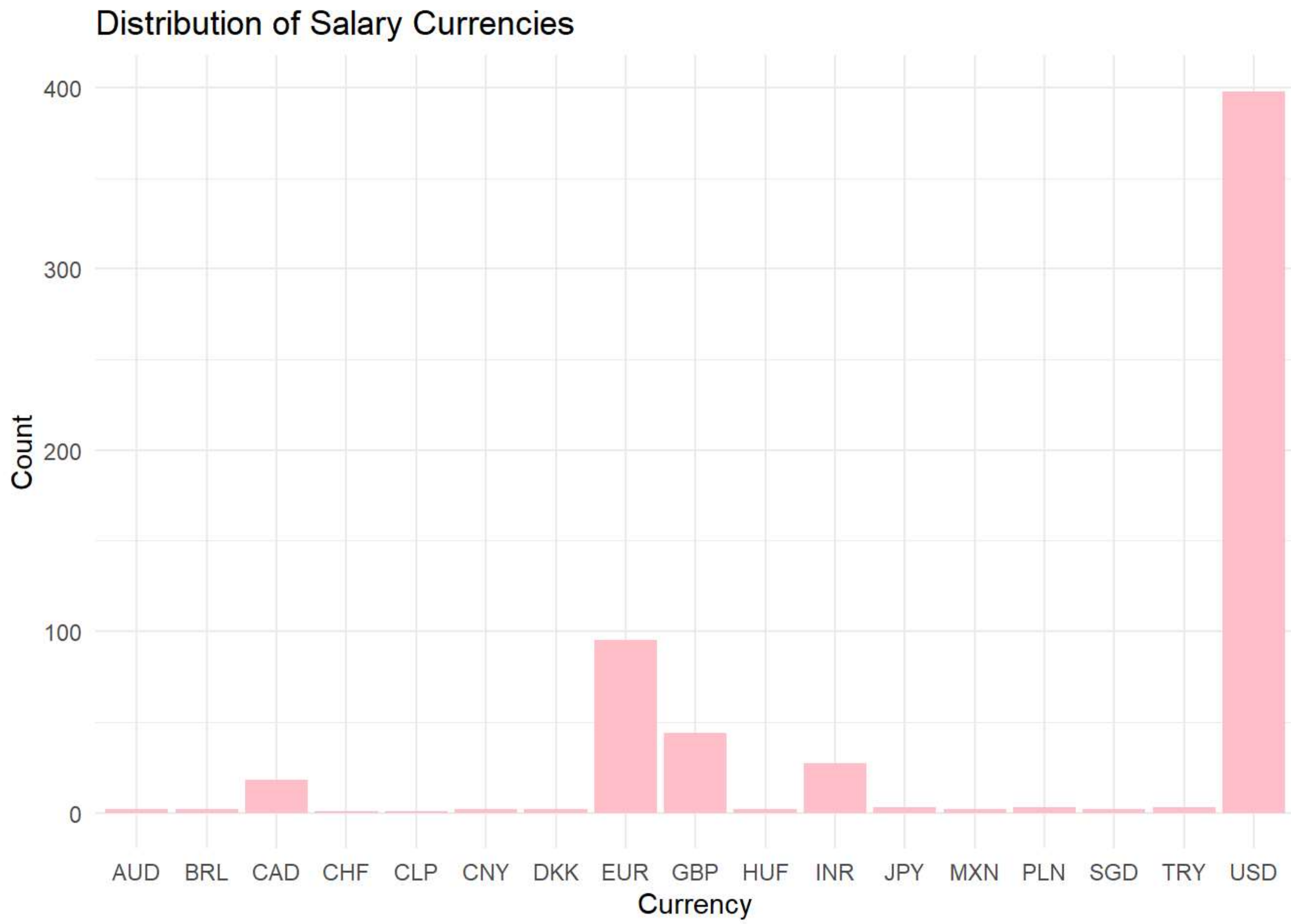

```
ggplot(data, aes(x = work_year)) +
  geom_bar(fill = "purple", alpha = 0.5) +
  labs(title = "Frequency of Work Year", x = "Work Year", y = "Count") +
  theme_minimal()
```




```
ggplot(data, aes(x = "", y = salary)) +
  geom_boxplot(fill = "orange", alpha = 0.7) +
  labs(title = "Salary Distribution", x = "", y = "Salary") +
  theme_minimal()
```

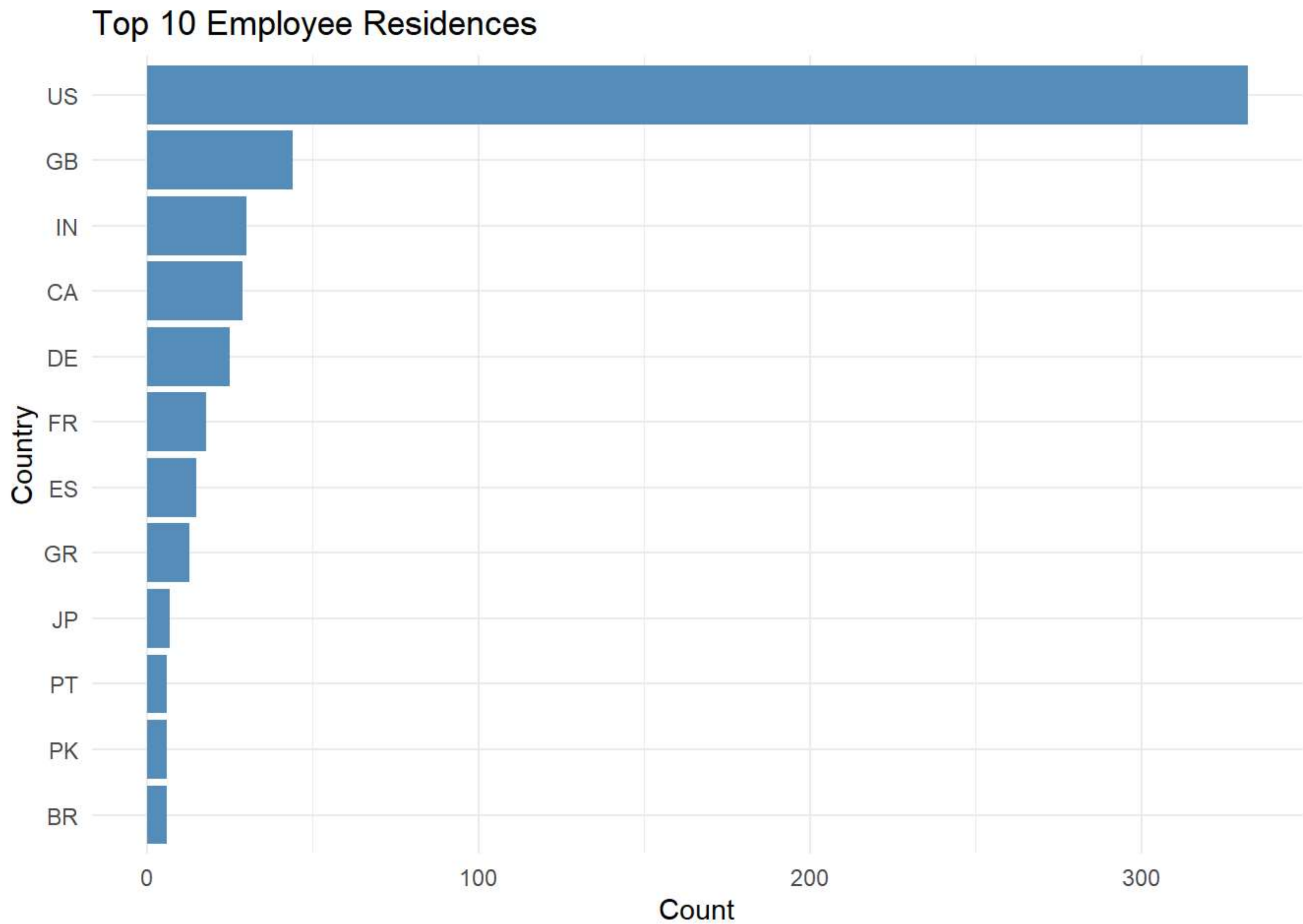



```
ggplot(data, aes(x = salary_currency)) +
  geom_bar(fill = "pink", alpha = 1.2) +
  labs(title = "Distribution of Salary Currencies", x = "Currency", y = "Count") +
  theme_minimal()
```



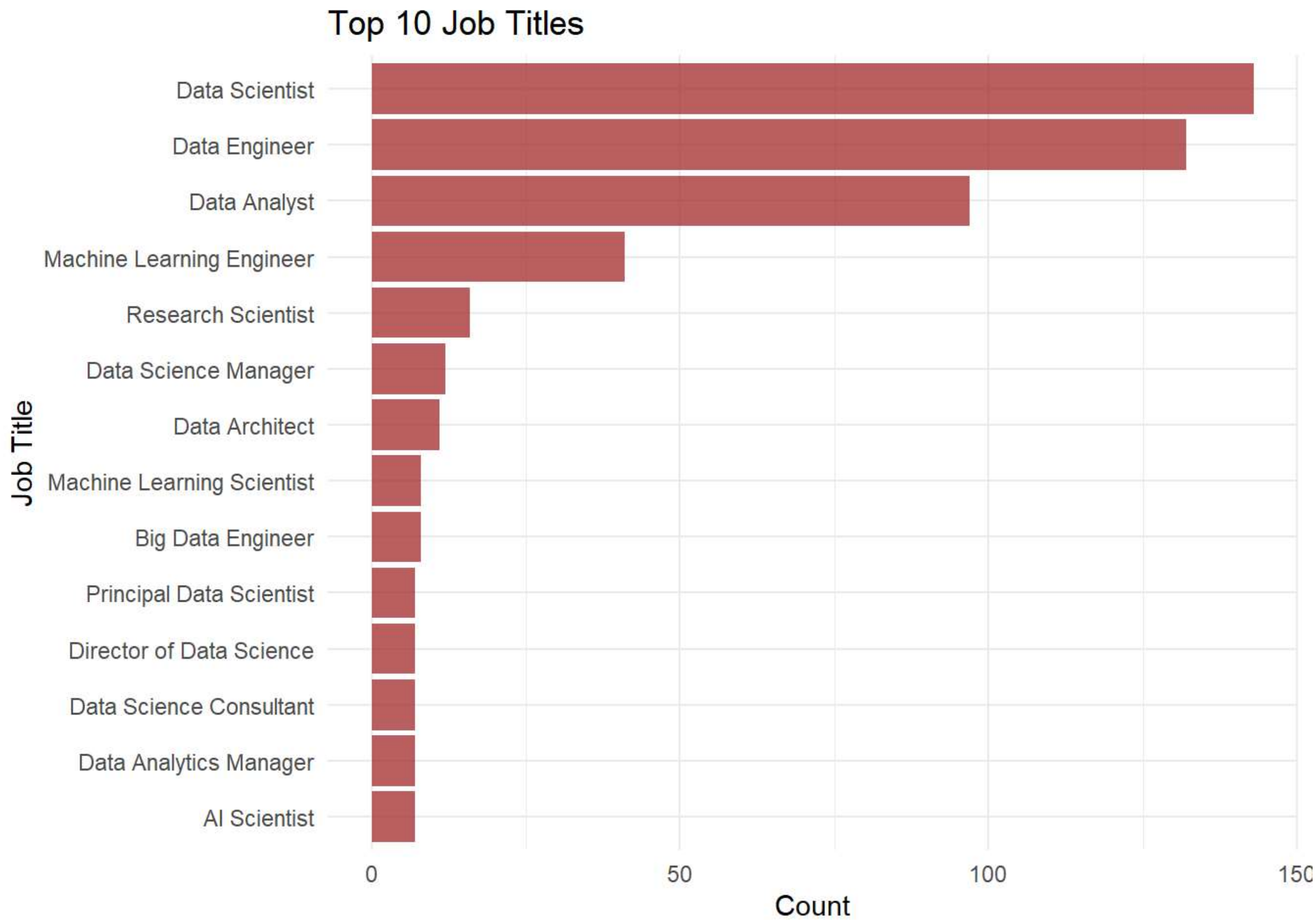

```
data %>%
  count(employee_residence, sort = TRUE) %>%
  top_n(10) %>%
  ggplot(aes(x = reorder(employee_residence, n), y = n)) +
  geom_bar(stat = "identity", fill = "steelblue", alpha = 0.9) +
  coord_flip() + #flip axis
  labs(title = "Top 10 Employee Residences", x = "Country", y = "Count") +
  theme_minimal()
```

```
## Selecting by n
```

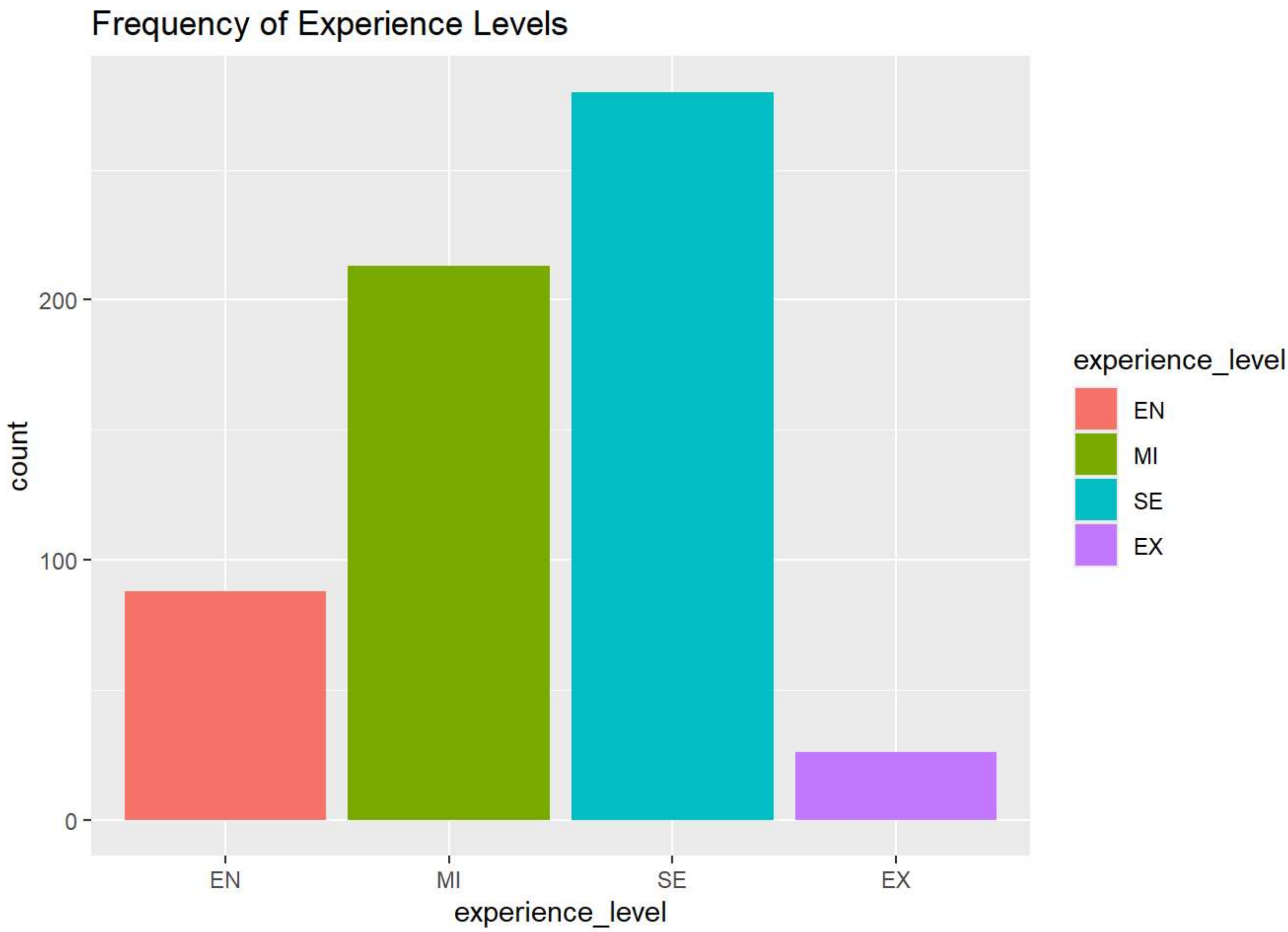



```
data %>%
  count(job_title, sort = TRUE) %>%
  top_n(10) %>%
  ggplot(aes(x = reorder(job_title, n), y = n)) +
  geom_bar(stat = "identity", fill = "brown", alpha = 0.75) +
  coord_flip() +
  labs(title = "Top 10 Job Titles", x = "Job Title", y = "Count") +
  theme_minimal()
```

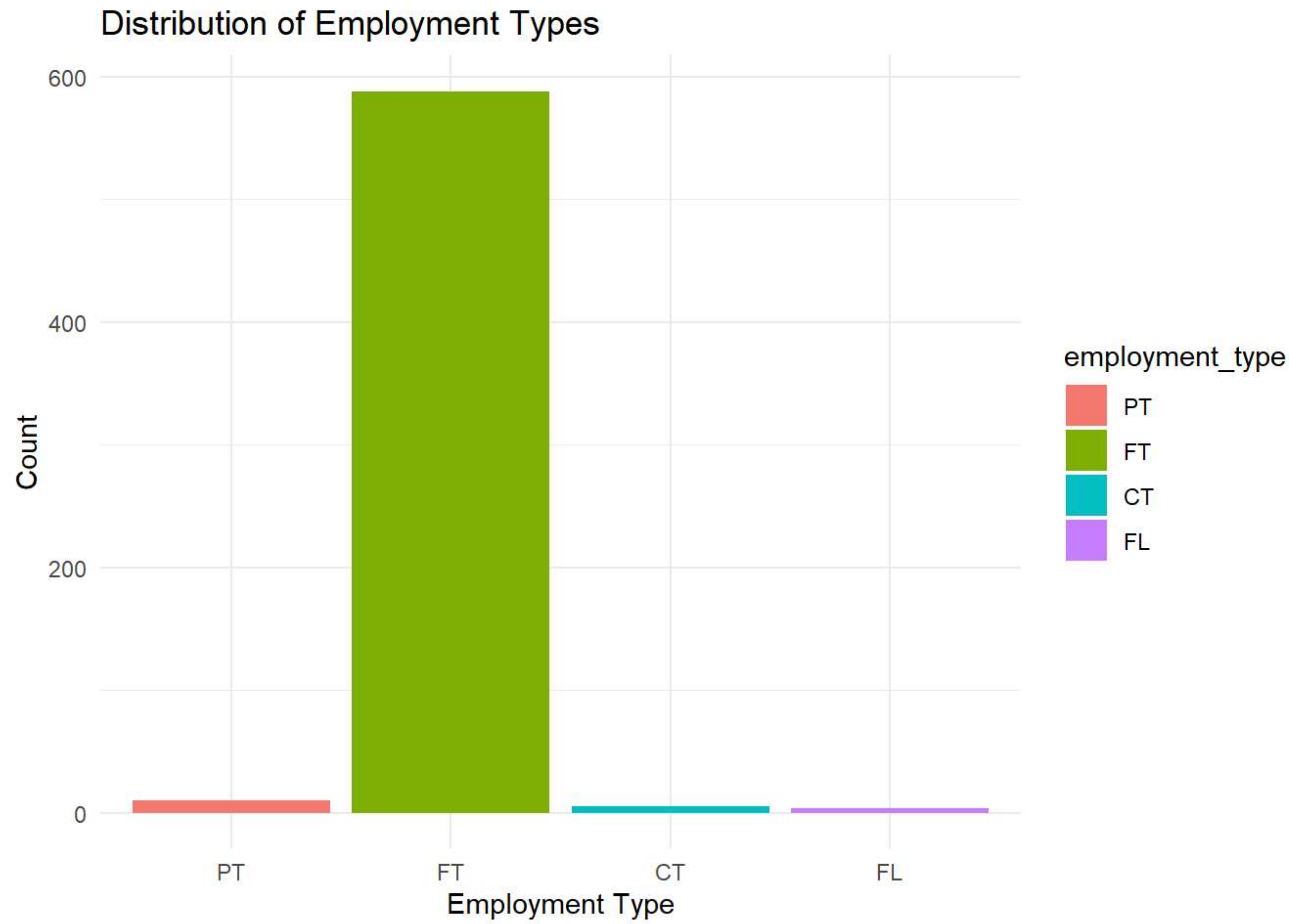
```
## Selecting by n
```




```
ggplot(data, aes(x = experience_level, fill = experience_level)) +
  geom_bar() +
  labs(title = "Frequency of Experience Levels")
```




```
data %>%
  count(employment_type) %>%
  ggplot(aes(x = employment_type, y = n, fill = employment_type)) +
  geom_col(alpha = 0.97) +
  labs(
    title = "Distribution of Employment Types",
    x = "Employment Type",
    y = "Count"
  ) +
  theme_minimal()
```




```
# Salary for Data related positions based on Experience Level and Employee Residence
summary_table <- data %>%
  group_by(experience_level, employee_residence) %>%
  summarise(
    mean_salary = mean(salary_in_usd, na.rm = TRUE),
    median_salary = median(salary_in_usd, na.rm = TRUE),
    count = n()
  )
```

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

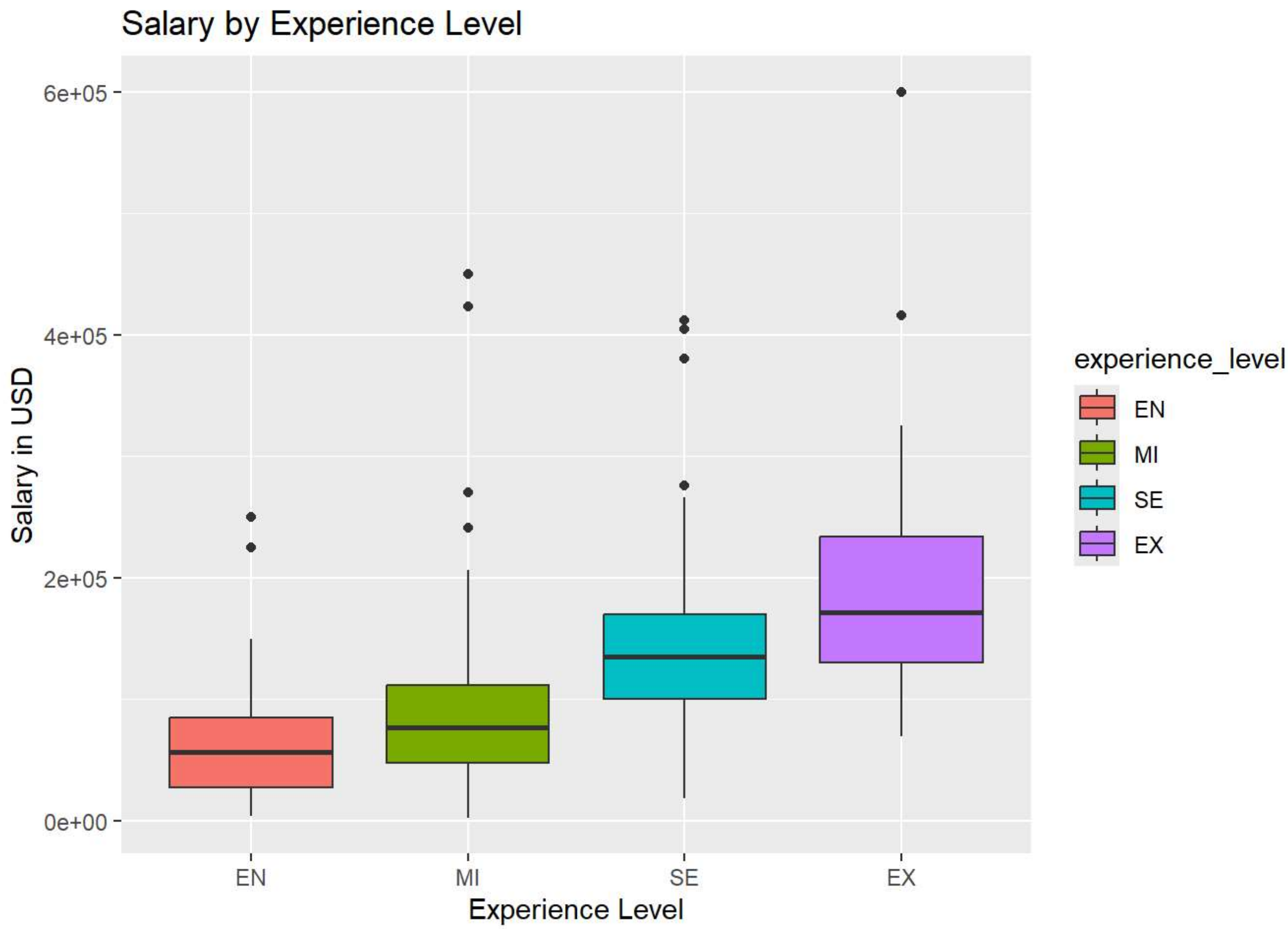
```
print(summary_table)
```

```
## # A tibble: 103 × 5
## # Groups:   experience_level [4]
##    experience_level employee_residence mean_salary median_salary count
##    <fct>            <chr>                    <dbl>         <dbl> <int>
##  1 EN               AU                     118352.       118352.     2
##  2 EN               BR                      12000         12000      1
##  3 EN               CA                      57132         52396      3
##  4 EN               CO                      21844         21844      1
##  5 EN               DE                      54412.        63278.     8
##  6 EN               DK                      37252.        37252.     2
##  7 EN               DZ                     100000        100000      1
##  8 EN               ES                      10354         10354      1
##  9 EN               FR                      47326.        49646      5
## 10 EN               GB                      65605.        52351      5
## # ℹ 93 more rows
```

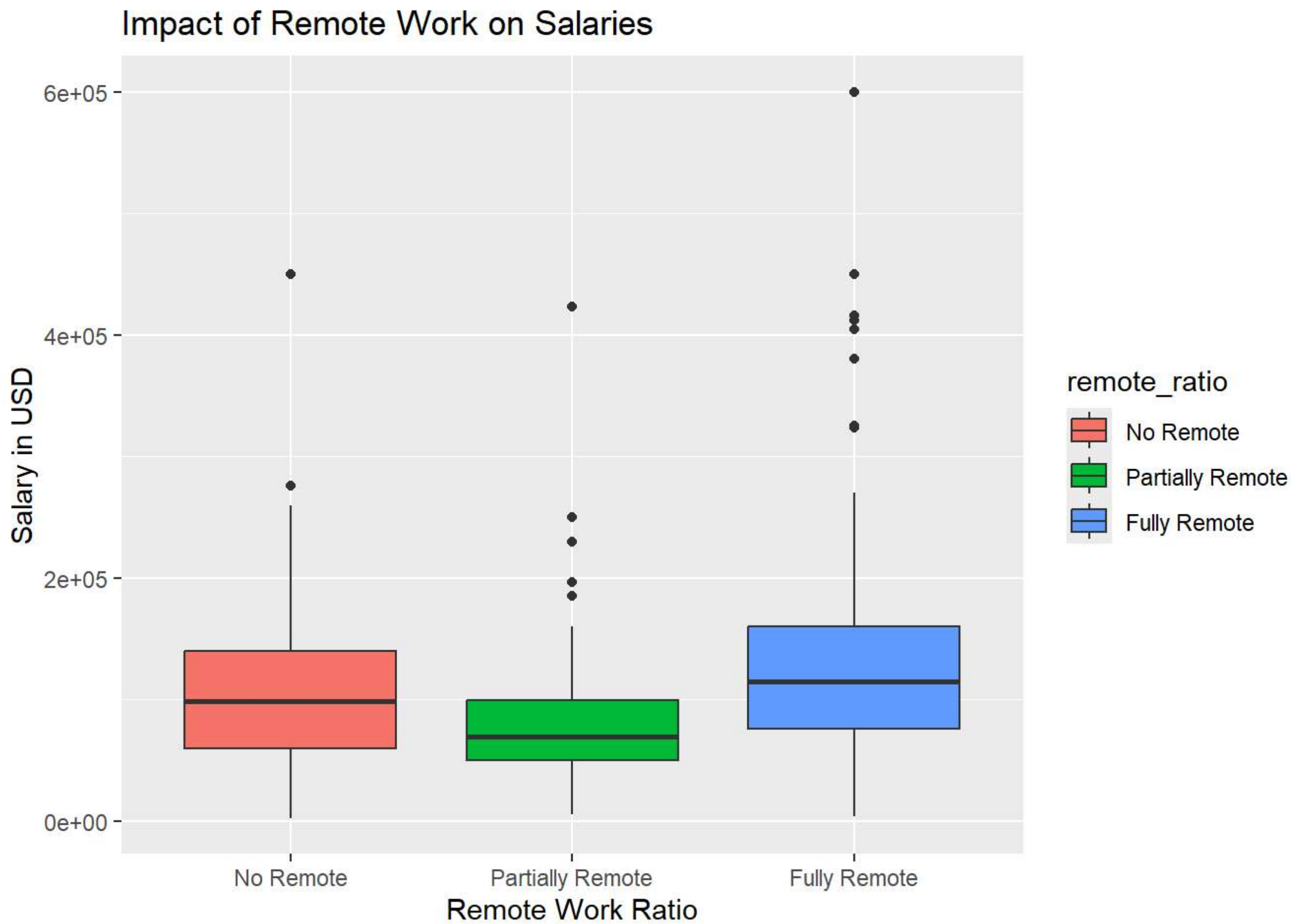
```
# U.S. versus Offshore Salary
comparison_table <- data %>%
  mutate(location_type = ifelse(employee_residence == "US", "US", "Offshore")) %>%
  group_by(location_type) %>%
  summarise(
    mean_salary = mean(salary_in_usd, na.rm = TRUE),
    median_salary = median(salary_in_usd, na.rm = TRUE),
    count = n()
  )
print(comparison_table)
```

```
## # A tibble: 2 × 4
##   location_type mean_salary median_salary count
##   <chr>               <dbl>         <dbl> <int>
## 1 Offshore           67754.         62649   275
## 2 US                149194.        138475   332
```

```
# Salary by Experience Level
ggplot(data, aes(x = experience_level, y = salary_in_usd, fill = experience_level)) +
  geom_boxplot() +
  labs(title = "Salary by Experience Level",
       y = "Salary in USD",
       x = "Experience Level")
```




```
# The impact of Remote Work on Salary
ggplot(data, aes(x = remote_ratio, y = salary_in_usd, fill = remote_ratio)) +
  geom_boxplot() +
  labs(title = "Impact of Remote Work on Salaries", y = "Salary in USD", x = "Remote Work Rati
o")
```



Conclusion:

We can see that the salary for data related positions is rather high in the US when compared to outsourcing data professionals from offshore countries. For example, the mean annual salary of a data related position in the US is 149,194 dollars compared to the offshore mean annual salary of 67,754 dollars. This equals a savings of roughly 81,000 dollars per employee if an offshore employee is hired instead. Therefore to limit costs spent on hiring data scientists, hiring highly experienced data professionals from offshore locations should be utilized. This is important because our company can hire well-equipped data professionals for the job and we can pay them less than we would if they were from the US, as their expectations for pay are lower. Another important thing to note is that the salaries of entry-level employees are generally lower than those of mid-level and senior employees based on the data. You may want to cut costs, however, sacrificing experience level may impact the company's overall efficiency. In this case, I believe that the best course of action would be to focus on hiring mid-level data professionals from offshore locations. This would allow the company to cut down on some costs while not overpaying a US employee with a significant amount of experience, and not sacrificing work quality by hiring only entry-level employees with minimal experience. The increase in salary on average from entry-level jobs to mid-level positions is much smaller than the jump from mid-level jobs to senior positions. Despite having slightly higher average annual salaries, senior level candidates are slightly less expensive as expert professionals. Finally, we must consider the impact of remote work on salaries. The average and the range of salaries for partially remote positions are the smallest, compared to the no remote work option salary which is slightly higher on average with a wider range, but the fully remote jobs have the highest average salary and generally have a higher range of salaries than partially remote and non-remote jobs. It makes most sense to hire a candidate that has mid-level to senior-level experience, works offshore, and as a result of working offshore that candidate works fully remote (relocation might require a higher salary).