R Notebook

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(here)
```

```
## here() starts at C:/Users/Dhruv/Desktop/Project_501/Project_501
```

```
library(readxl)
library(ggridges)
```

```
##
## Attaching package: 'ggridges'

## The following object is masked from 'package:ggplot2':
##
##     scale_discrete_manual
```

```
library(ggplot2)
library(gganimate)
library(gifski)
library(png)
```

```
bank_additional <- read_excel("bank-additional.xlsx")
bank_additional
```

```
## # A tibble: 4,119 x 21
##      age job   marital education default housing loan  contact month
##    <dbl> <chr> <chr>   <chr>     <chr>   <chr>   <chr> <chr>   <chr>
##  1    30 blue~ married basic.9y  no      yes     no    cellul~ may
##  2    39 serv~ single  high.sch~ no      no      no    teleph~ may
##  3    25 serv~ married high.sch~ no      yes     no    teleph~ jun
##  4    38 serv~ married basic.9y  no      unknown unkn~ teleph~ jun
##  5    47 admi~ married universi~ no      yes     no    cellul~ nov
##  6    32 serv~ single  universi~ no      no      no    cellul~ sep
##  7    32 admi~ single  universi~ no      yes     no    cellul~ sep
##  8    41 entr~ married universi~ unknown yes     no    cellul~ nov
```

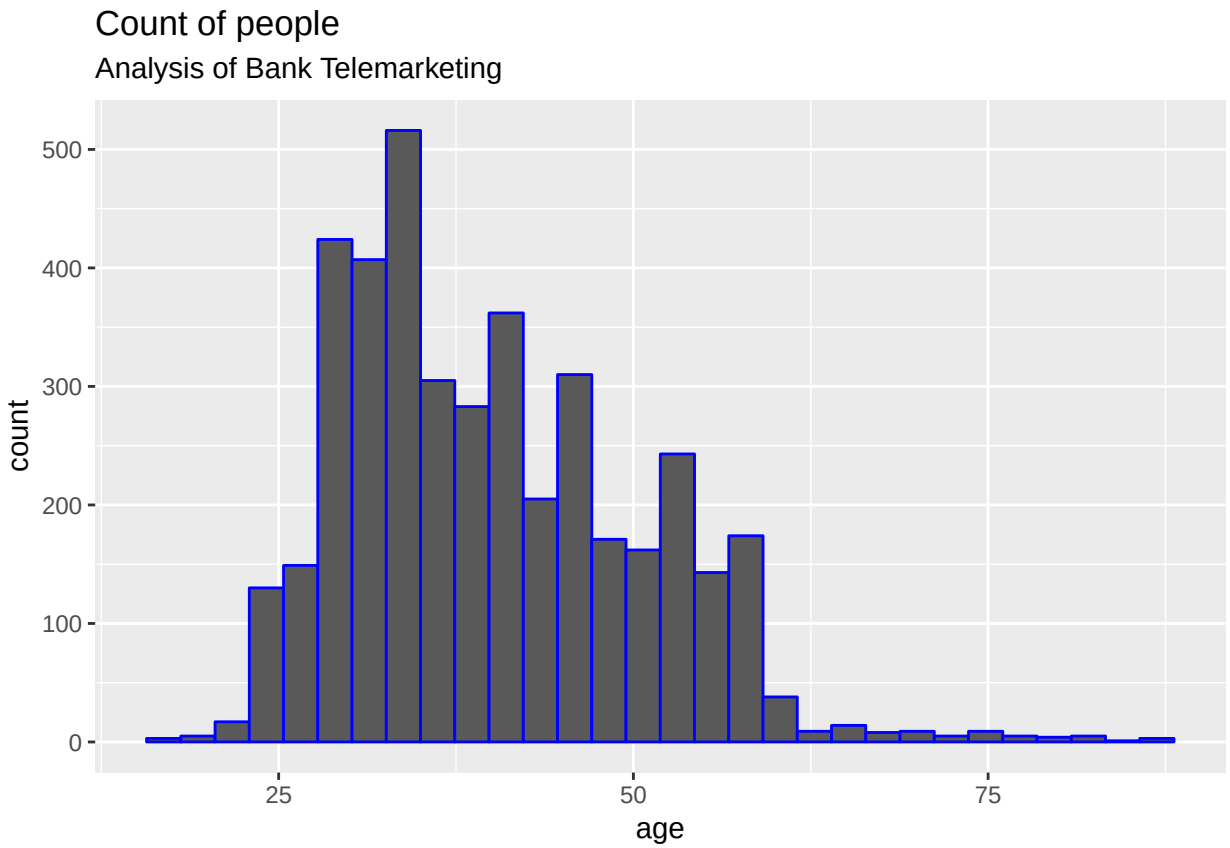
```
##  9    31 serv~ divorc~ professi~ no      no      no    cellul~ nov
## 10    35 blue~ married basic.9y  unknown no      no    teleph~ may
## # ... with 4,109 more rows, and 12 more variables: day_of_week <chr>,
## #   duration <dbl>, campaign <dbl>, pdays <dbl>, previous <dbl>,
## #   poutcome <chr>, emp.var.rate <dbl>, cons.price.idx <dbl>,
## #   cons.conf.idx <dbl>, euribor3m <dbl>, nr.employed <dbl>,
## #   term_deposit <chr>
```

```
glimpse(bank_additional)
```

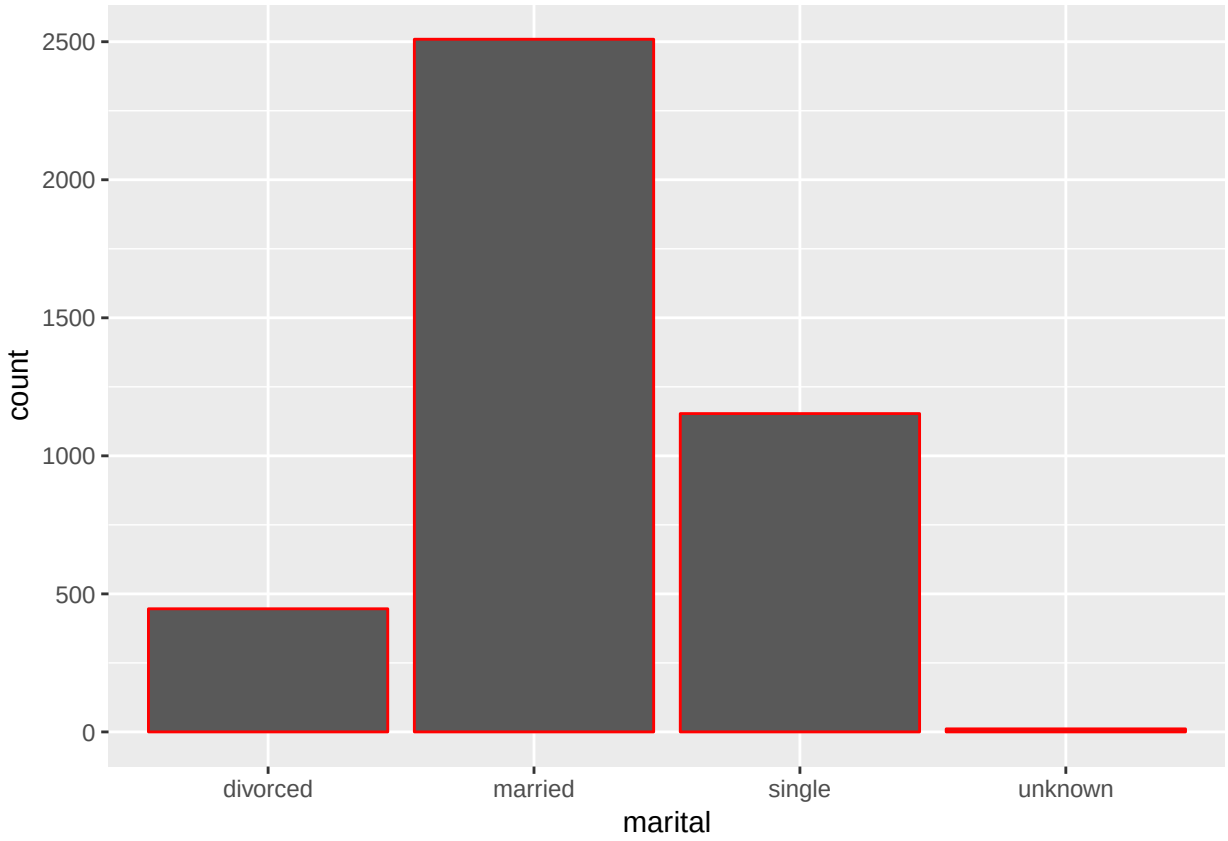
```
## Observations: 4,119
## Variables: 21
## $ age            <dbl> 30, 39, 25, 38, 47, 32, 32, 41, 31, 35, 25, 36,...
## $ job            <chr> "blue-collar", "services", "services", "service...
## $ marital        <chr> "married", "single", "married", "married", "mar...
## $ education      <chr> "basic.9y", "high.school", "high.school", "basi...
## $ default        <chr> "no", "no", "no", "no", "no", "no", "no", "unkn...
## $ housing        <chr> "yes", "no", "yes", "unknown", "yes", "no", "ye...
## $ loan           <chr> "no", "no", "no", "unknown", "no", "no", "no", ...
## $ contact        <chr> "cellular", "telephone", "telephone", "telephon...
## $ month          <chr> "may", "may", "jun", "jun", "nov", "sep", "sep"...
## $ day_of_week    <chr> "fri", "fri", "wed", "fri", "mon", "thu", "mon"...
## $ duration       <dbl> 487, 346, 227, 17, 58, 128, 290, 44, 68, 170, 3...
## $ campaign       <dbl> 2, 4, 1, 3, 1, 3, 4, 2, 1, 1, 1, 1, 2, 2, 2, 2,...
## $ pdays          <dbl> 999, 999, 999, 999, 999, 999, 999, 999, 999, 99...
## $ previous       <dbl> 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,...
## $ poutcome       <chr> "nonexistent", "nonexistent", "nonexistent", "n...
## $ emp.var.rate   <dbl> -1.8, 1.1, 1.4, 1.4, -0.1, -1.1, -1.1, -0.1, -0...
## $ cons.price.idx <dbl> 92.893, 93.994, 94.465, 94.465, 93.200, 94.199,...
## $ cons.conf.idx  <dbl> -46.2, -36.4, -41.8, -41.8, -42.0, -37.5, -37.5...
## $ euribor3m      <dbl> 1.313, 4.855, 4.962, 4.959, 4.191, 0.884, 0.879...
## $ nr.employed    <dbl> 5099.1, 5191.0, 5228.1, 5228.1, 5195.8, 4963.6,...
## $ term_deposit   <chr> "no", "no", "no", "no", "no", "no", "no", "no",...
```

```
ggplot(bank_additional) +
geom_histogram(bank_additional,mapping = aes(x = age), color = 'blue') +
labs(title="Count of people", subtitle="Analysis of Bank Telemarketing")
```

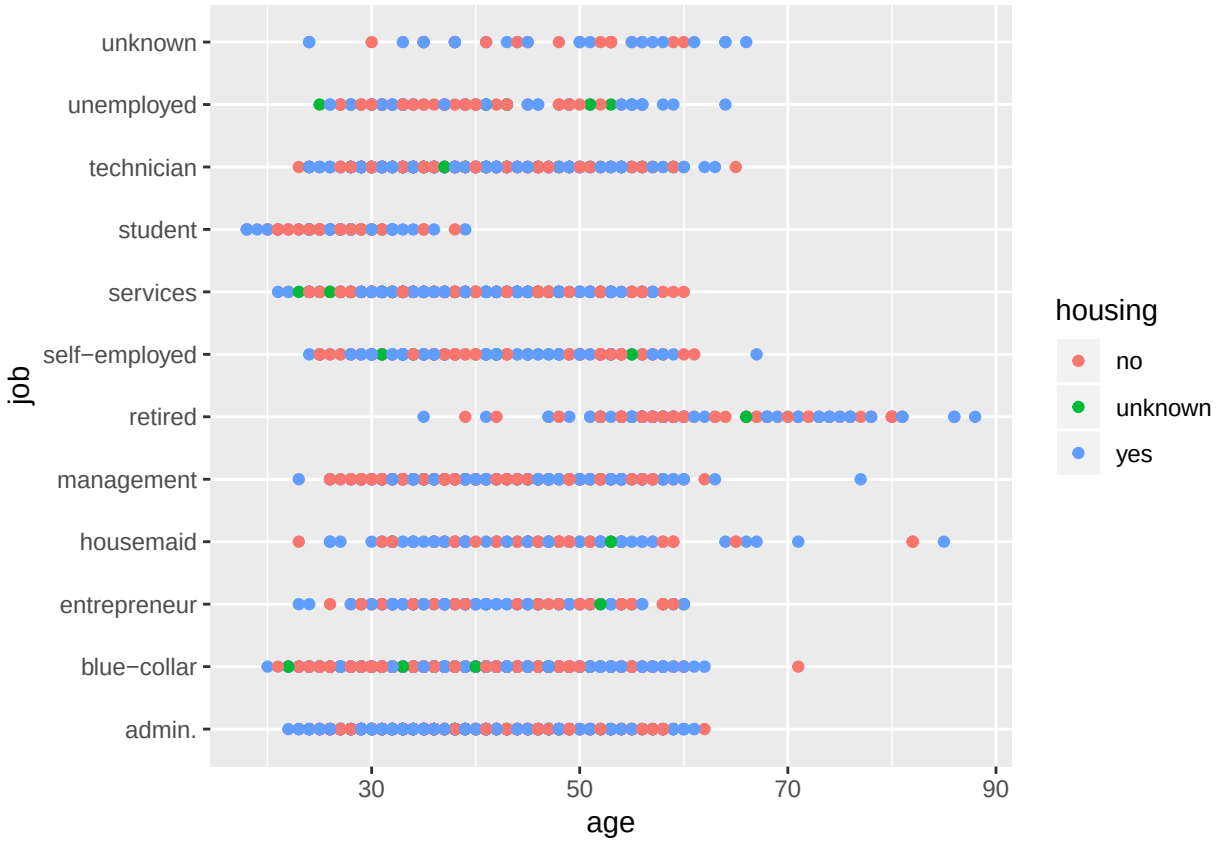
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(bank_additional) +
geom_bar(bank_additional,mapping = aes(x = marital),color = 'red')
```

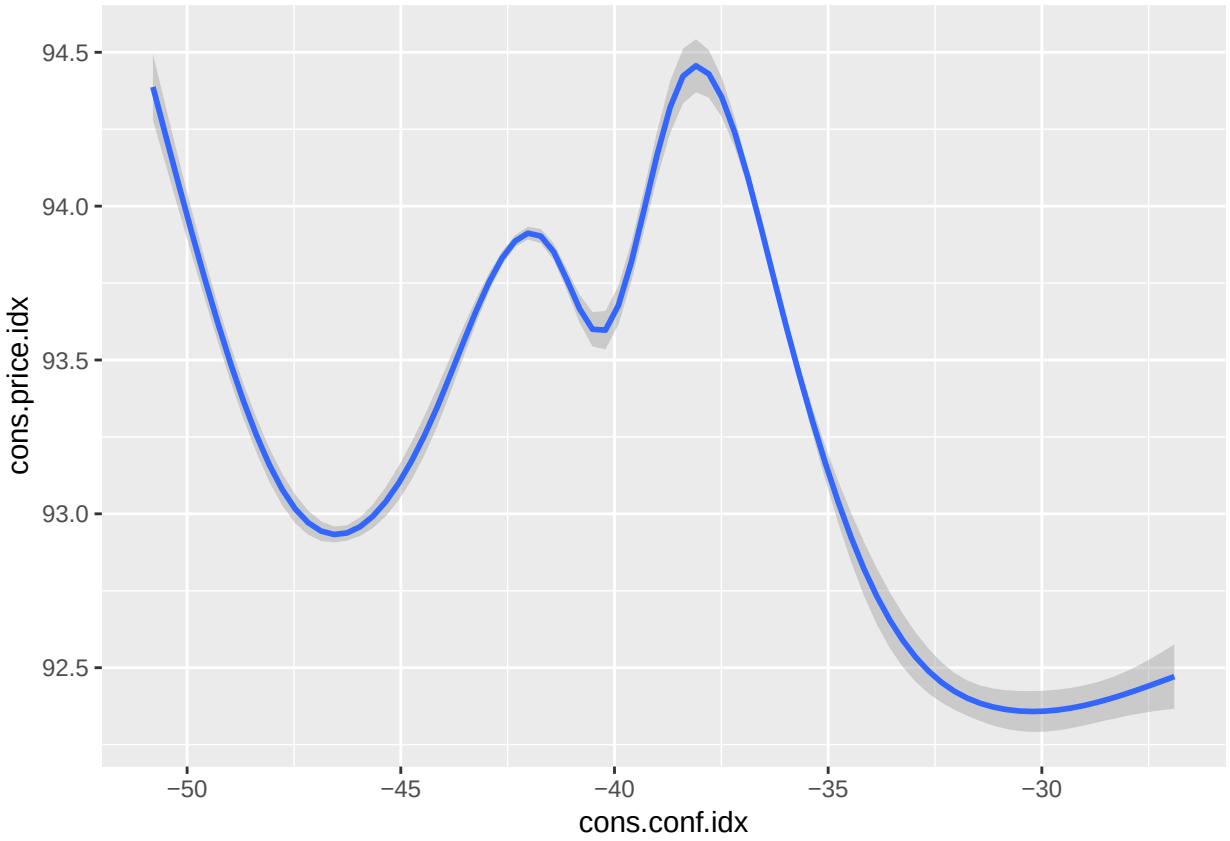



```
ggplot(data = bank_additional) +
geom_point(bank_additional,mapping = aes(x = age, y = job, color = housing))
```

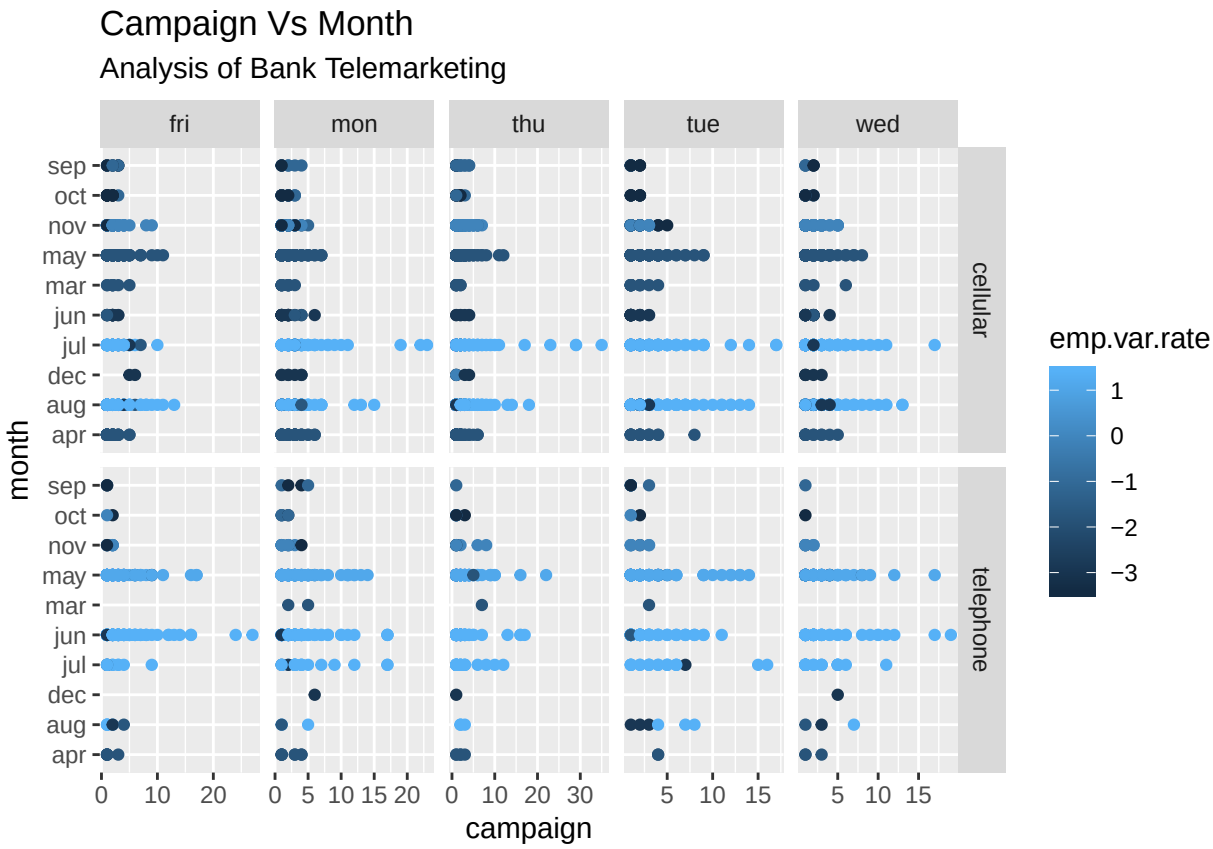



```
ggplot(bank_additional,mapping = aes(x = cons.conf.idx, y = cons.price.idx)) +
geom_smooth()
```

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




```
ggplot(bank_additional) +
geom_point(mapping = aes(x = campaign, y = month, color = emp.var.rate)) +
facet_grid(contact ~ day_of_week,scales = 'free')  +
labs(title="Campaign Vs Month", subtitle="Analysis of Bank Telemarketing")
```


Campaign Vs Month
Analysis of Bank Telemarketing
fri
mon
thu
tue
wed
cellular
telephone
month
sep
oct
nov
may
mar
jun
jul
dec
aug
apr
0 10 20
0 5 10 15 20
0 10 20 30
5 10 15
5 10 15
campaign
emp.var.rate
1
0
−1
−2
−3