EmployeeReview-Vivek

SaiNagaChandraVivekGarimella

August 14, 2019

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

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

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v tibble  1.4.2     v purrr   0.2.5
## v tidyr   0.8.2     v stringr 1.3.1
## v tibble  1.4.2     v forcats 0.3.0
```

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

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
read_data <- read_csv("C:/Users/garim/OneDrive/Desktop/NEU/Introduction to Data Management and Processi
                  na = c("none", "n/a") )
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   company = col_character(),
##   location = col_character(),
##   dates = col_character(),
##   `job-title` = col_character(),
##   summary = col_character(),
##   pros = col_character(),
##   cons = col_character(),
##   `advice-to-mgmt` = col_character(),
```

```
##   `overall-ratings` = col_double(),
##   `work-balance-stars` = col_double(),
##   `culture-values-stars` = col_double(),
##   `carrer-opportunities-stars` = col_double(),
##   `comp-benefit-stars` = col_double(),
##   `senior-mangemnet-stars` = col_double(),
##   `helpful-count` = col_double(),
##   link = col_character()
## )
```

```
locations_df <- read.csv("file:///C:/Users/garim/OneDrive/Desktop/NEU/Introduction to Data Management an



tidy_data <- read_data %>% full_join(locations_df) %>%
            select(-X) %>%
            separate(`job-title` , into=c("job-title","designation"), sep = "-")
```

```
## Joining, by = "location"
```

```
## Warning: Column `location` joining character vector and factor, coercing
## into character vector
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 642 rows [146,
## 242, 252, 827, 954, 1314, 1892, 1997, 2594, 2670, 4649, 5135, 5331, 5716,
## 6006, 6138, 6179, 6343, 6595, 6693, ...].
```

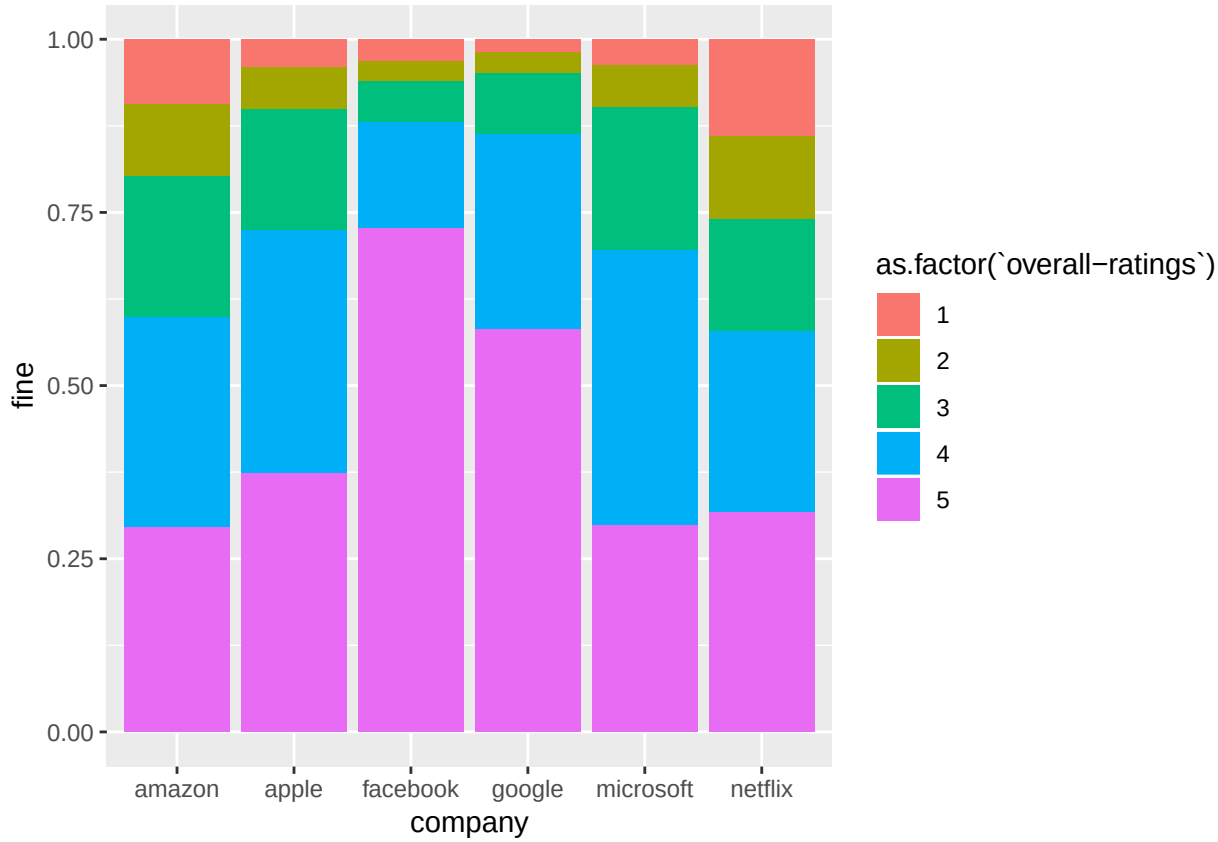
```
general_data<- read_data

 general_data<-general_data %>% filter(`overall-ratings` !="none" & `culture-values-stars`!="none" & `wo

 companynames <-general_data %>% group_by(company) %>% summarise(observations = n())
company_overallrating <-general_data %>% group_by(`overall-ratings`,company) %>% summarise(rating = n())

# overal rating proportions each company
full_join(companynames,company_overallrating) %>% mutate(fine=rating/observations) %>% ggplot()+geom_bar
```

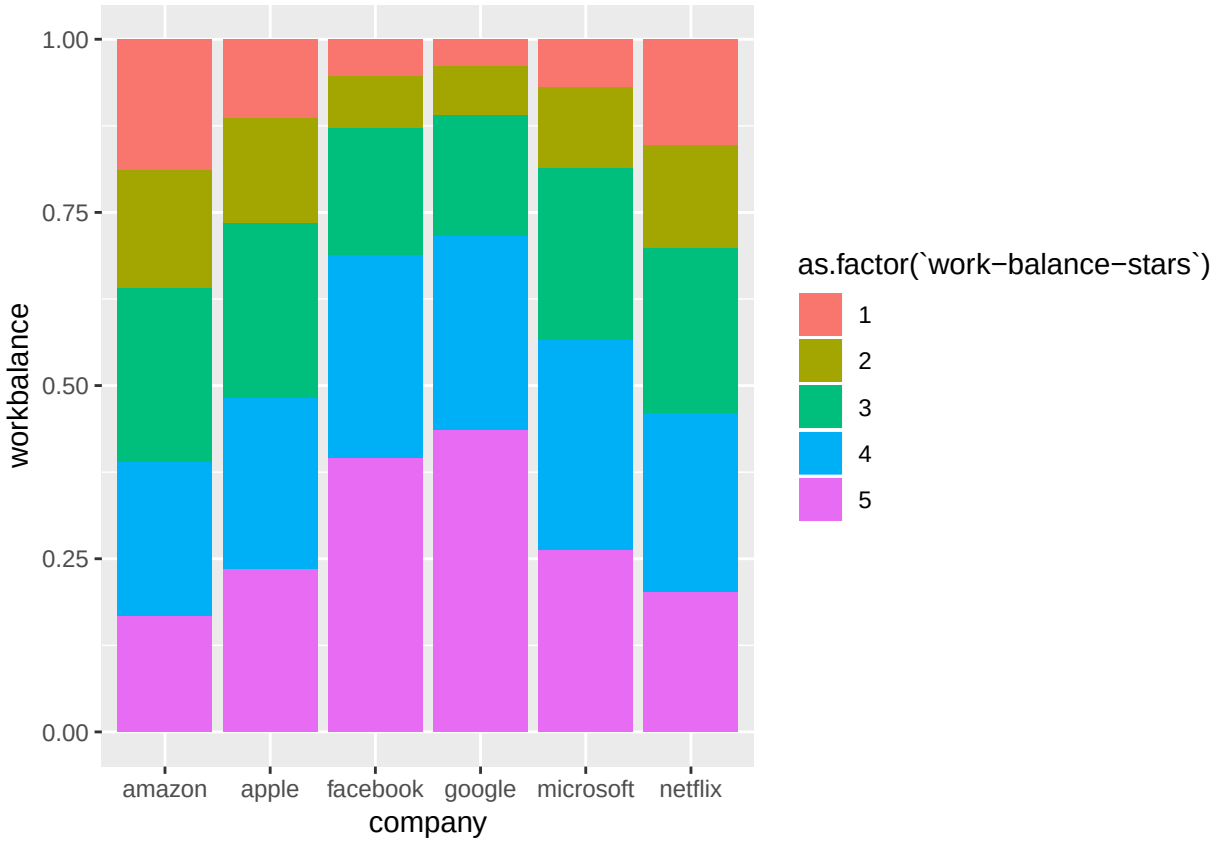
```
## Joining, by = "company"
```




```
company_workbalance<- general_data  %>% group_by(`work-balance-stars`,company) %>% summarise(total=n())


full_join(company_workbalance,companynames) %>% mutate(workbalance= total/observations ) %>% ggplot()+g
```

```
## Joining, by = "company"
```




```
general_data[["dates"]]<-mdy(general_data[["dates"]])
```

```
## Warning: 4 failed to parse.
```

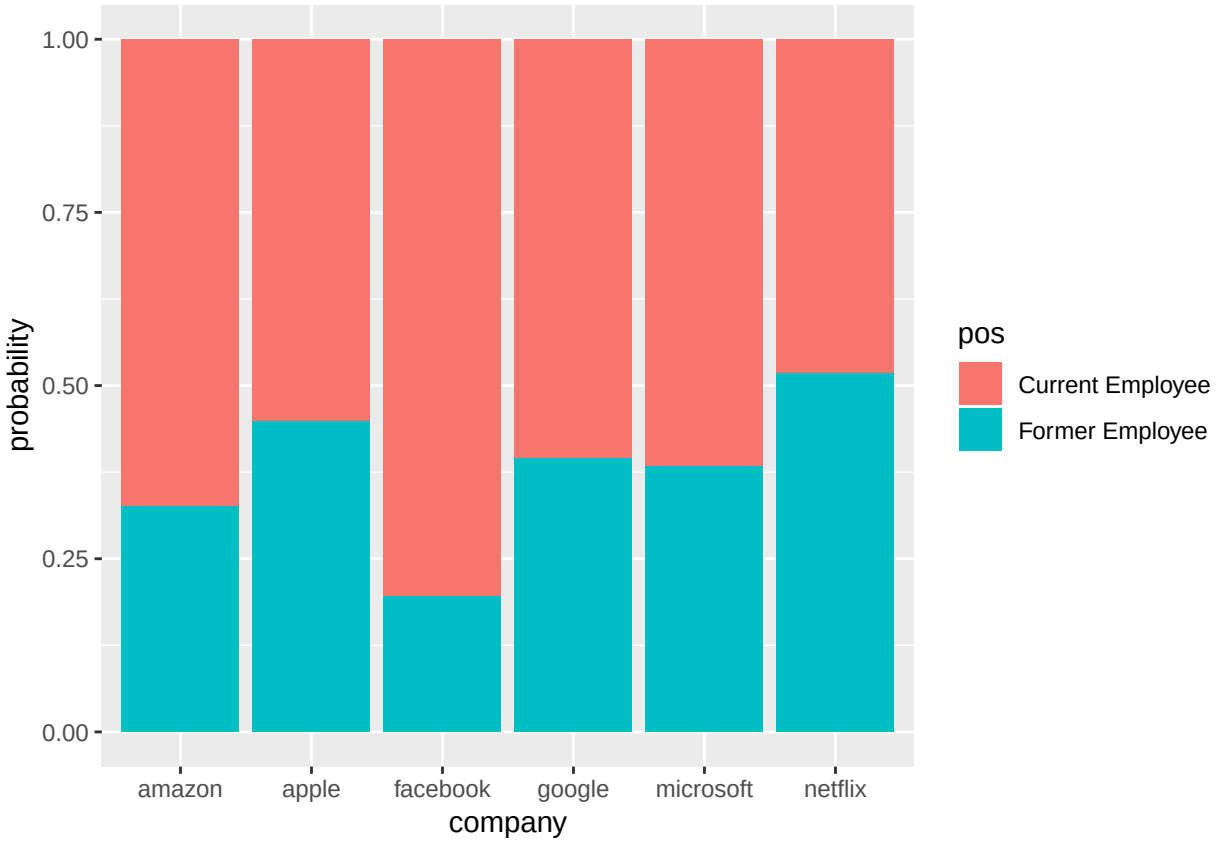
```
general_data<-mutate(general_data,year=year(dates))

general_data<-mutate(general_data,pos= str_sub(general_data[["job-title"]],1,16))
general_data<-mutate(general_data,pos=str_trim(general_data[["pos"]]))


company_position <-general_data %>% group_by(pos,company) %>% summarise(position=n())

 full_join(company_position,companynames) %>% mutate(probability=position/observations) %>% ggplot()+ge
```

```
## Joining, by = "company"
```



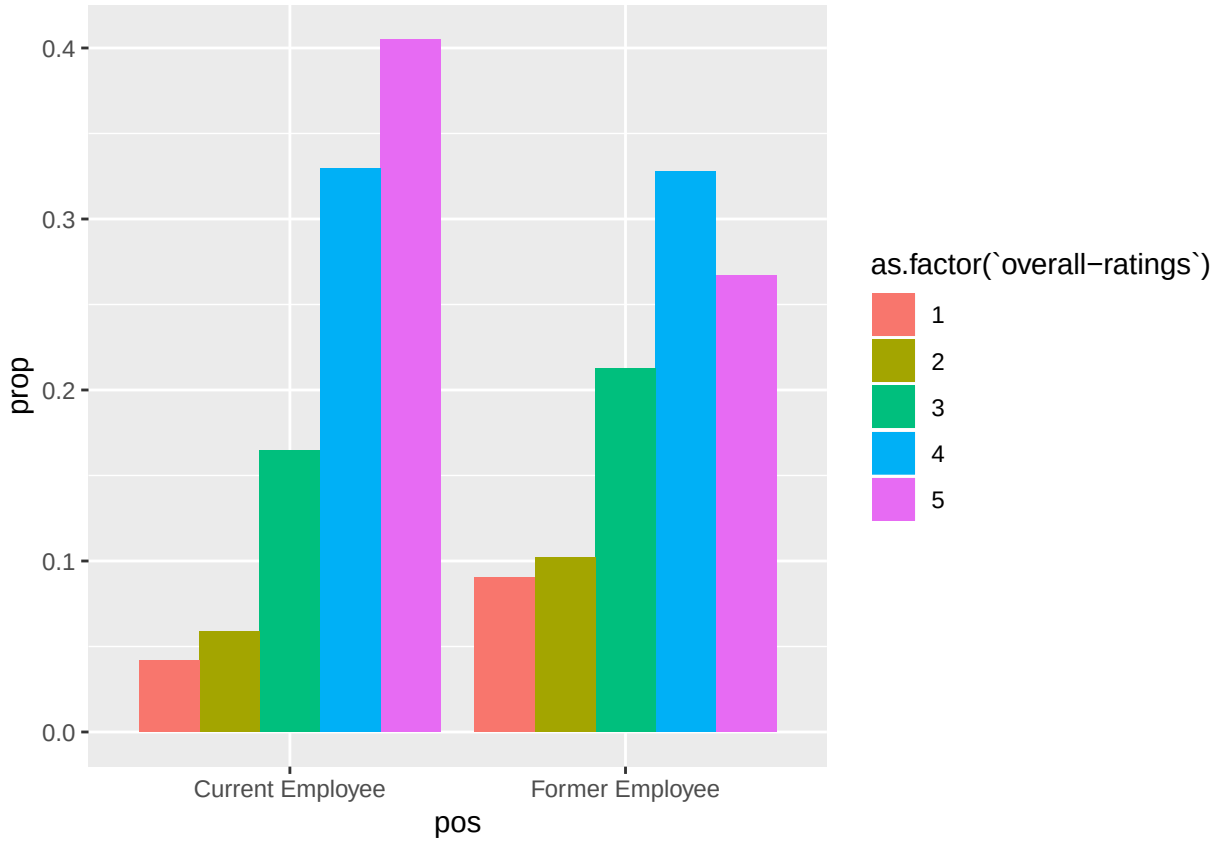

```
general_data
```

```
## # A tibble: 53,222 x 19
##       X1 company location dates      `job-title` summary pros  cons
##    <dbl> <chr>   <chr>    <date>     <chr>       <chr>   <chr> <chr>
##  1     1 google  <NA>     2018-12-11 Current Em~ Best C~ Peop~ Bure~
##  2     2 google  Mountai~ 2013-06-21 Former Emp~ Moving~ "1) ~ "1) ~
##  3     3 google  New Yor~ 2014-05-10 Current Em~ Great ~ "* I~ * It~
##  4     4 google  Mountai~ 2015-02-08 Current Em~ The be~ You ~ I li~
##  5     5 google  Los Ang~ 2018-07-19 Former Emp~ Unique~ Goog~ "If ~
##  6     6 google  Mountai~ 2018-12-09 Former Emp~ NICE w~ Peop~ Food~
##  7     7 google  New Yor~ 2018-12-11 Current Em~ Softwa~ Grea~ Usua~
##  8     8 google  <NA>     2018-12-11 Former Emp~ great ~ work~ No c~
##  9     9 google  New Yor~ 2018-12-10 Current Em~ Google~ Grea~ Youn~
## 10    10 google  <NA>     2018-12-09 Current Em~ Execel~ Impa~ Size~
## # ... with 53,212 more rows, and 11 more variables:
## #   `advice-to-mgmt` <chr>, `overall-ratings` <dbl>,
## #   `work-balance-stars` <dbl>, `culture-values-stars` <dbl>,
## #   `carrer-opportunities-stars` <dbl>, `comp-benefit-stars` <dbl>,
## #   `senior-mangemnet-stars` <dbl>, `helpful-count` <dbl>, link <chr>,
## #   year <dbl>, pos <chr>
```

```
positions<-general_data %>% group_by(pos) %>% summarise(totalpos=n())
overall_pos<-general_data %>% group_by(pos,`overall-ratings`) %>% summarise(eachpos=n())

full_join(positions,overall_pos) %>% mutate(prop=eachpos/totalpos) %>% ggplot()+geom_bar(mapping = aes(
```

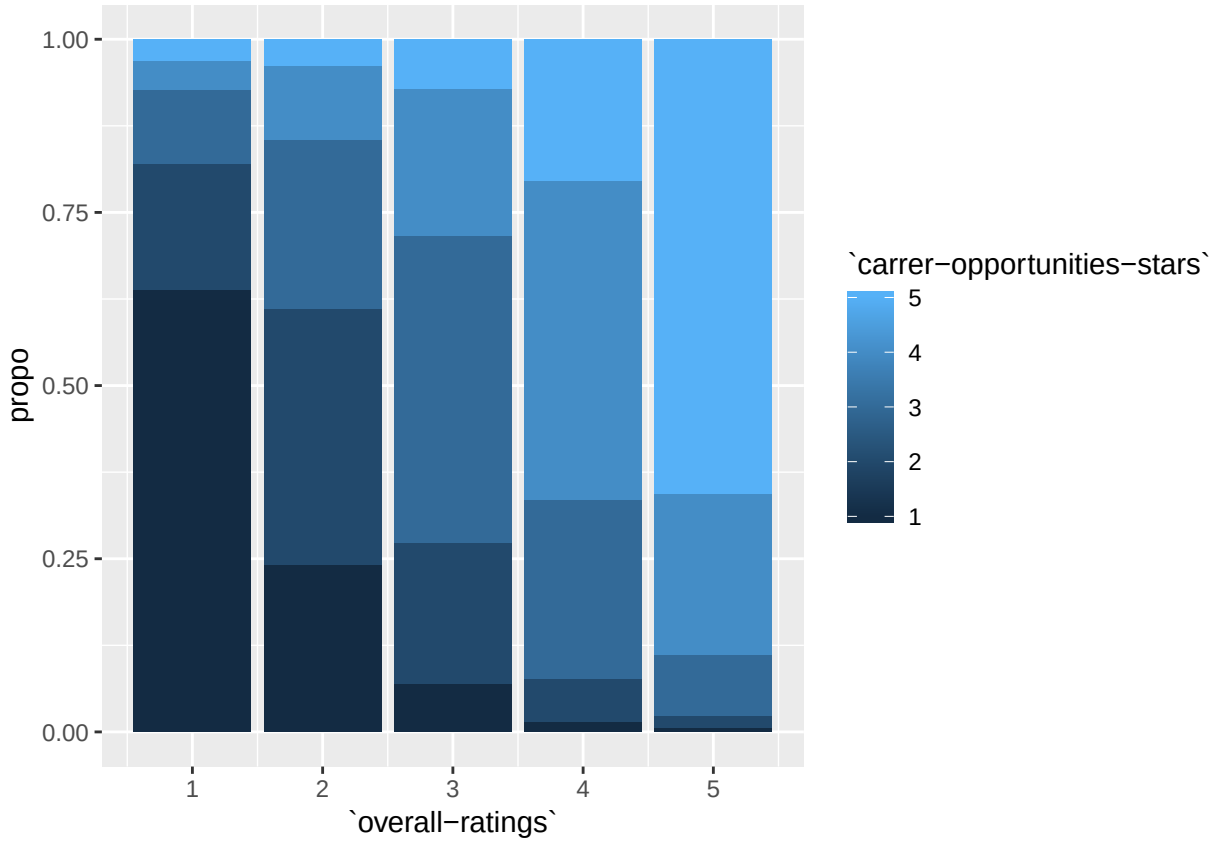
```
## Joining, by = "pos"
```




```
overall_ratings<-general_data %>% group_by(`overall-ratings`) %>% summarise(overall=n())
overall_ratings_career<-general_data %>% group_by(`overall-ratings`,`carrer-opportunities-stars`) %>% su

full_join(overall_ratings,overall_ratings_career) %>% mutate(propo=overallcareer/overall) %>% ggplot()+g
```

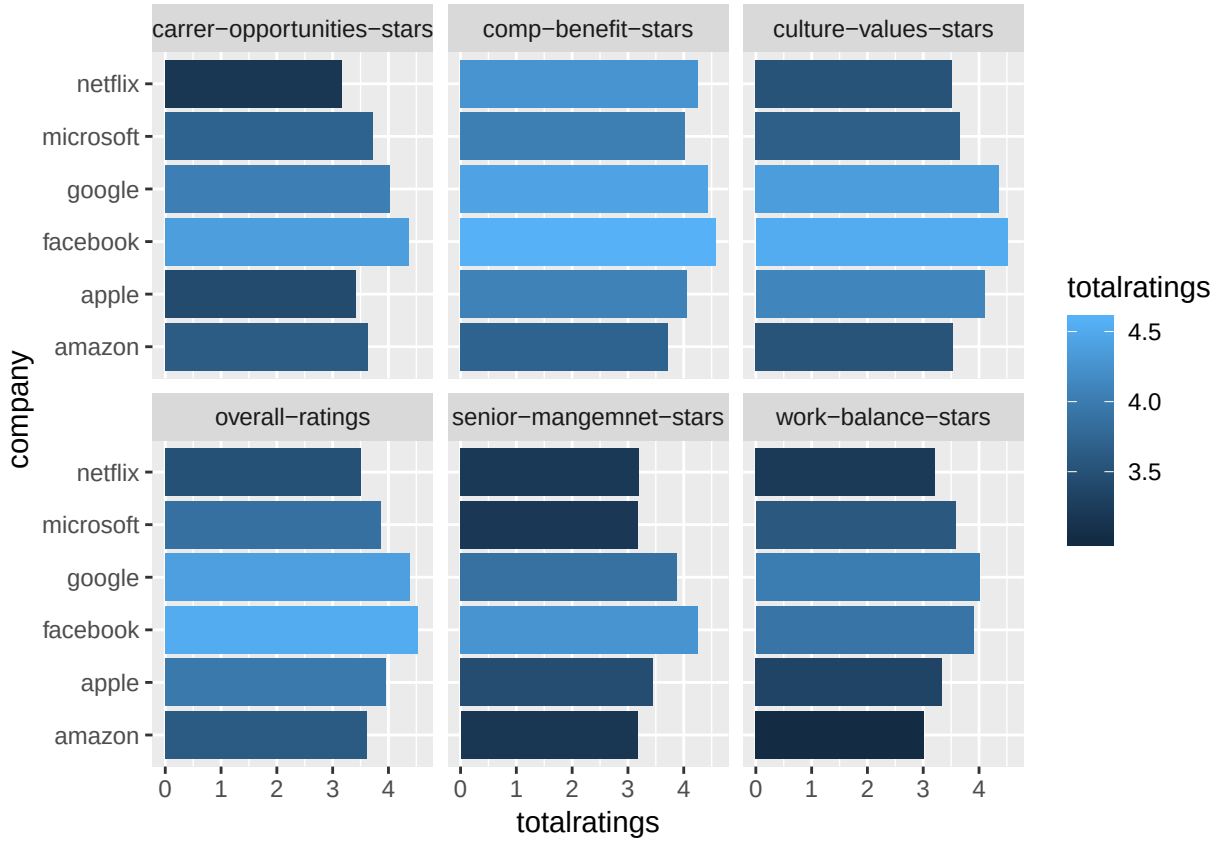
```
## Joining, by = "overall-ratings"
```




```
average<- general_data %>% gather(`overall-ratings`,`work-balance-stars`,`culture-values-stars`,`carrer
```

```
full_join(average,companynames) %>% group_by(company,fields) %>% mutate(totalratings=totalsum/observati
```

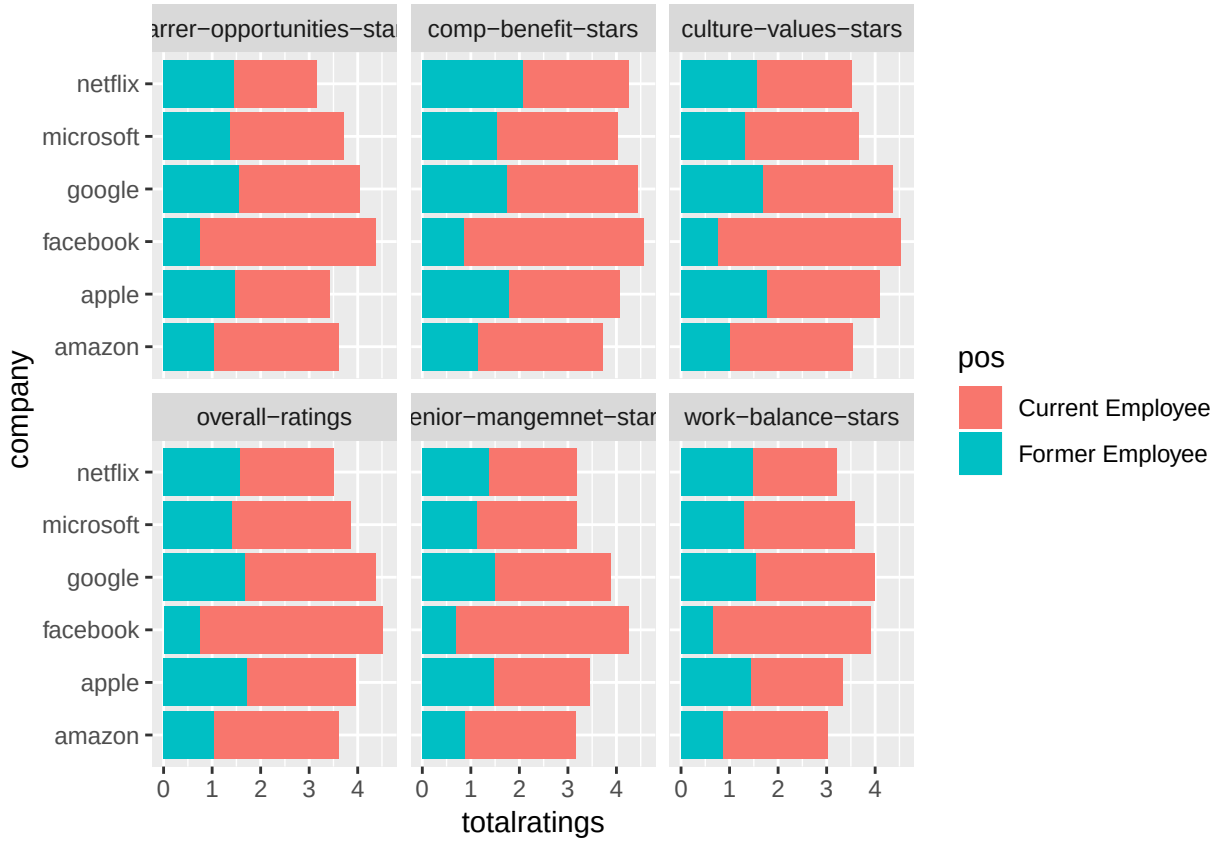
```
## Joining, by = "company"
```




```
average_pos<- general_data %>% gather(`overall-ratings`,`work-balance-stars`,`culture-values-stars`,`ca

full_join(companynames,average_pos)  %>% mutate(totalratings=numerator/observations) %>% ggplot()+geom_
```

```
## Joining, by = "company"
```


arrer−opportunities−sta
comp−benefit−stars
culture−values−stars
overall−ratings
enior−mangemnet−star
work−balance−stars
company
netflix
microsoft
google
facebook
apple
amazon
0
1
2
3
4
totalratings
pos
Current Employee
Former Employee