At work on Friday, I was trying to figure out the best way to display some rank data. What I had were rankings from 1-5 for 10 factors considered most important in a job (such as Salary, Insurance Benefits, and the Opportunity to Learn), meaning each respondent chose and ranked the top 5 from those 10, and the remaining 5 were unranked by that respondent. Without even thinking about the missing data issue, I computed a mean rank and called it a day. (Yes, I know that ranks are ordinal and means are for continuous data, but my goal was simply to differentiate importance of the factors and a mean seemed the best way to do it.) Of course, then we noticed one of the factors had a pretty high average rank, even though few people ranked it in the top 5. Oops.

So how could I present these results? One idea I had was a stacked bar chart, and it took a bit of data wrangling to do it. That is, the rankings were all in separate variables, but I want them all on the same chart. Basically, I needed to create a dataset with:

1 variable to represent the factor being ranked

* 1 variable to represent the ranking given (1-5, or 6 that I called “Not Ranked”)
* 1 variable to represent the number of people giving that particular rank that particular factor

What I ultimately did was run frequencies for the factor variables, turn those frequency tables into data frames, and merged them together with rbind. I then created chart with ggplot. Here’s some code for a simplified example, which only uses 6 factors and asks people to rank the top 3.

First, let’s read in our [sample dataset](https://www.dropbox.com/s/93g4mrm1rbrj5tl/sample_ranks.csv?dl=0) – note that these data were generated only for this example and are not real data:

library(tidyverse)

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

## v ggplot2 3.0.0 v purrr 0.2.4  
## v tibble 1.4.2 v dplyr 0.7.4  
## v tidyr 0.8.0 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## Warning: package 'ggplot2' was built under R version 3.5.1

## -- Conflicts ------------------------------------------------------------------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

ranks <- read\_csv("C:/Users/slocatelli/Desktop/sample\_ranks.csv", col\_names = TRUE)

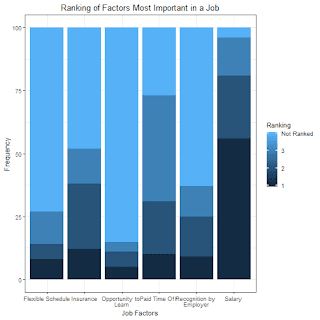
## Parsed with column specification:  
## cols(  
## RespID = col\_integer(),  
## Salary = col\_integer(),  
## Recognition = col\_integer(),  
## PTO = col\_integer(),  
## Insurance = col\_integer(),  
## FlexibleHours = col\_integer(),  
## OptoLearn = col\_integer()  
## )

This dataset contains 7 variables – 1 respondent ID and 6 variables with ranks on factors considered important in a job: salary, recognition from employer, paid time off, insurance benefits, flexible scheduling, and opportunity to learn. I want to run frequencies for these variables, and turn those frequency tables into a data frame I can use in ggplot2. I’m sure there are much cleaner ways to do this (and please share in the comments!), but here’s one not so pretty way:

salary <- as.data.frame(table(ranks$Salary))  
salary$Name <- "Salary"  
recognition <- as.data.frame(table(ranks$Recognition))  
recognition$Name <- "Recognition by \nEmployer"  
PTO <- as.data.frame(table(ranks$PTO))  
PTO$Name <- "Paid Time Off"  
insurance <- as.data.frame(table(ranks$Insurance))  
insurance$Name <- "Insurance"  
flexible <- as.data.frame(table(ranks$FlexibleHours))  
flexible$Name <- "Flexible Schedule"  
learn <- as.data.frame(table(ranks$OptoLearn))  
learn$Name <- "Opportunity to \nLearn"  
  
rank\_chart <- rbind(salary, recognition, PTO, insurance, flexible, learn)  
rank\_chart$Var1 <- as.numeric(rank\_chart$Var1)

With my not-so-pretty data wrangling, the chart itself is actually pretty easy:

ggplot(rank\_chart, aes(fill = Var1, y = Freq, x = Name)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Ranking of Factors Most Important in a Job") +  
 ylab("Frequency") +  
 xlab("Job Factors") +  
 scale\_fill\_continuous(name = "Ranking",  
 breaks = c(1:4),  
 labels = c("1","2","3","Not Ranked")) +  
 theme\_bw() +  
 theme(plot.title=element\_text(hjust=0.5))

[](https://i1.wp.com/3.bp.blogspot.com/-QTzTFhtVuLs/XEuE4ahJv2I/AAAAAAAAPHg/A5Bu3an8180GrCl5ZhkFu_nqkxbnUjIIwCLcBGAs/s1600/unnamed-chunk-3-1.png?ssl=1)

Based on this chart, we can see the top factor is Salary. Insurance is slightly more important than paid time off, but these are definitely the top 2 and 3 factors. Recognition wasn’t ranked by most people, but those who did considered it their #2 factor; ditto for flexible scheduling at #3. Opportunity to learn didn’t make the top 3 for most respondents.