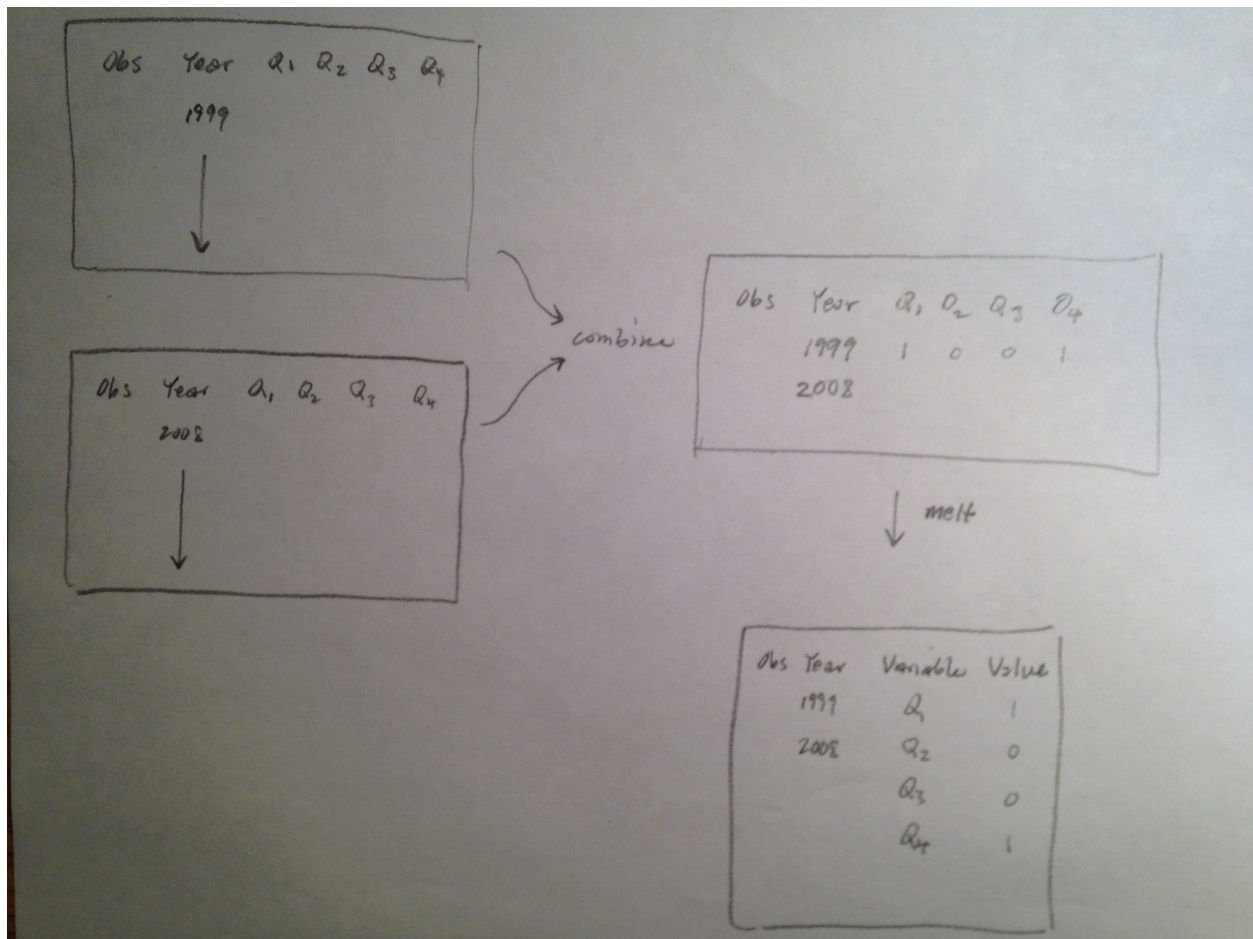# Homework 3

Chen, Frank fc2451

October 14, 2014

```r
library("foreign")
library("plyr")
library("ggplot2")
library("ggvis")
library("reshape2")
setwd("/Users/knarf/Copy/datasets/columbia/4_dv/")
gesis.1999 <- read.dta("ZA3777_v3-0-1.dta")
gesis.2008 <- read.dta("ZA4752_v1-0-0.dta")
g99.job <- gesis.1999[,83:100]
g08.job <- gesis.2008[,81:99]
```

Introduction

The purpose of this assignment is to compare the responses of two similar survey questions from the European Values survey database. The corresponding question from 1999 (question 13) and 2008 (question 14) ask Great Britianers which aspects of a job they think are especially significant. The respondents would give their answer, and their answers would be compared to approximately 20 groups of themes (e.g "good pay", "pleasant people to work with", "not too much pressure", "good job security". . . etc). Each theme would be labeled with a binary classification of "mentioned" or "not mentioned."

The graph that I would like to make would be a series of paired histogram or paired point comparisons. In addition to these static histograms, a histogram with a slider, utilizing ggplot and ggvis, where one could observe the differences in aggregated frequency response counts for the question between the two time periods of each survey can also be made.

Obs Year Q₁ Q₂ Q₃ Q₄
1999

Obs Year Q₁ Q₂ Q₃ Q₄
2008

combine

Obs Year Q₁ Q₂ Q₃ Q₄
1999 1 0 0 1
2008

melt

Obs Year	Variable	Value
1999	Q₁	1
2008	Q₂	0
	Q₃	0
	Q₄	1

Functions

```r
hate.factors <- function(x) { # turn factors into characters
    x <- sapply(x, as.character)
    return(x)
}

love.numbers <- function(x) { # changes characters to binary 1/0
    x[x == "mentioned"] <- 1
    x[x == "not mentioned"] <- 0
    x[x == "job aspect mentioned"] <- 1
    x[x == "no job aspect mentioned"] <- 0
    x[x == "consistent"] <- 1
    return(x)
}

evict.na <- function(x) { # handling missing data in 2008
    x[x == "inconsistent 7"] <- 0
    x[x == "inconsistent 6"] <- 0
    x[x == "inconsistent 1"] <- 0
    x[, ncol(x)][is.na(x[, ncol(x)])] <- 0
```

```r
    x <- x[complete.cases(x),]
    return(x)
}

proportionize <- function(x) { # proportion by year
    rows <- nrow(x)
    x <- colSums(x)
    x <- x/rows
    x <- t(as.data.frame(x))
    return(x)
}

name.1999 <- function(x) { # question naming
    colnames(x) <- c("date", "good pay", "pleasant co-workers", "less pressure", "good security", "prome
    return(x)
}

name.2008 <- function(x) { # question naming
    colnames(x) <- c("date", "good pay", "pleasant co-workers", "less pressure", "good security", "good
    return(x)
}
```

Processing & Melting

```r
g99.job <- hate.factors(g99.job)
g99.job <- love.numbers(g99.job)
g99.job <- as.matrix(as.numeric(g99.job))
dim(g99.job) <- c(1000, 18)
g99.job <- proportionize(g99.job)
g99.date <- matrix(1999, nrow=1, ncol=1)
g99.job <- cbind(g99.date, g99.job)
g99.job <- name.1999(g99.job)
g99.job <- as.data.frame(g99.job)

g08.job <- hate.factors(g08.job)
g08.job <- love.numbers(g08.job)
g08.job <- evict.na(g08.job)
g08.job <- as.matrix(as.numeric(g08.job))
dim(g08.job) <- c(1533, 19)
g08.job <- proportionize(g08.job)
g08.date <- matrix(2008, nrow=1, ncol=1)
g08.job <- cbind(g08.date, g08.job)
g08.job <- name.2008(g08.job)
g08.job <- as.data.frame(g08.job)

agg.val <- merge(g08.job, g99.job, all.x = TRUE, all.y = TRUE)
agg.val[, 1:ncol(agg.val)][is.na(agg.val[, 1:ncol(agg.val)])] <- 0
agg.val <- melt(agg.val, id.var = "date")
agg.val
```

```
##    date         variable      value
```

```
## 1  1999          good pay 0.8130000
## 2  2008          good pay 0.7390737
## 3  1999 pleasant co-workers 0.7180000
## 4  2008 pleasant co-workers 0.7110241
## 5  1999      less pressure 0.2900000
## 6  2008      less pressure 0.2380952
## 7  1999      good security 0.6570000
## 8  2008      good security 0.6549250
## 9  1999         good hours 0.5720000
## 10 2008         good hours 0.5231572
## 11 1999     uses initiative 0.4110000
## 12 2008     uses initiative 0.4468363
## 13 1999            useful 0.2830000
## 14 2008            useful 0.3542074
## 15 1999   generous holidays 0.3820000
## 16 2008   generous holidays 0.2961513
## 17 1999      meeting people 0.4300000
## 18 2008      meeting people 0.4181344
## 19 1999 feeling achievement 0.5960000
## 20 2008 feeling achievement 0.6581866
## 21 1999     responsibility 0.3720000
## 22 2008     responsibility 0.3913894
## 23 1999        interesting 0.6810000
## 24 2008        interesting 0.6914547
## 25 1999      uses abilities 0.3980000
## 26 2008      uses abilities 0.4324853
## 27 1999         job aspect 0.9980000
## 28 2008         job aspect 0.9732551
## 29 1999     learning skills 0.0000000
## 30 2008     learning skills 0.4794521
## 31 1999     family friendly 0.0000000
## 32 2008     family friendly 0.3378995
## 33 1999     leadership role 0.0000000
## 34 2008     leadership role 0.3052838
## 35 1999     equal treatment 0.0000000
## 36 2008     equal treatment 0.6477495
## 37 1999        consistency 0.0000000
## 38 2008        consistency 0.9804305
## 39 1999    promotion chance 0.3700000
## 40 2008    promotion chance 0.0000000
## 41 1999            respect 0.2600000
## 42 2008            respect 0.0000000
## 43 1999    good conditions 0.4670000
## 44 2008    good conditions 0.0000000
## 45 1999        weekends off 0.4380000
## 46 2008        weekends off 0.0000000
```

Plots

```
ggplot(agg.val, aes(variable, value, fill=as.factor(date))) +
    geom_bar(data = agg.val, position="dodge", stat="identity") +
```

```
    coord_flip() +
    ylab("Proportion") +
    xlab("Survey Question") +
    ggtitle("Important Job Aspects in Great Britain") +
    scale_fill_manual(values = c("red", "blue")) +
    theme_bw() #+ facet_wrap(~variable)
```
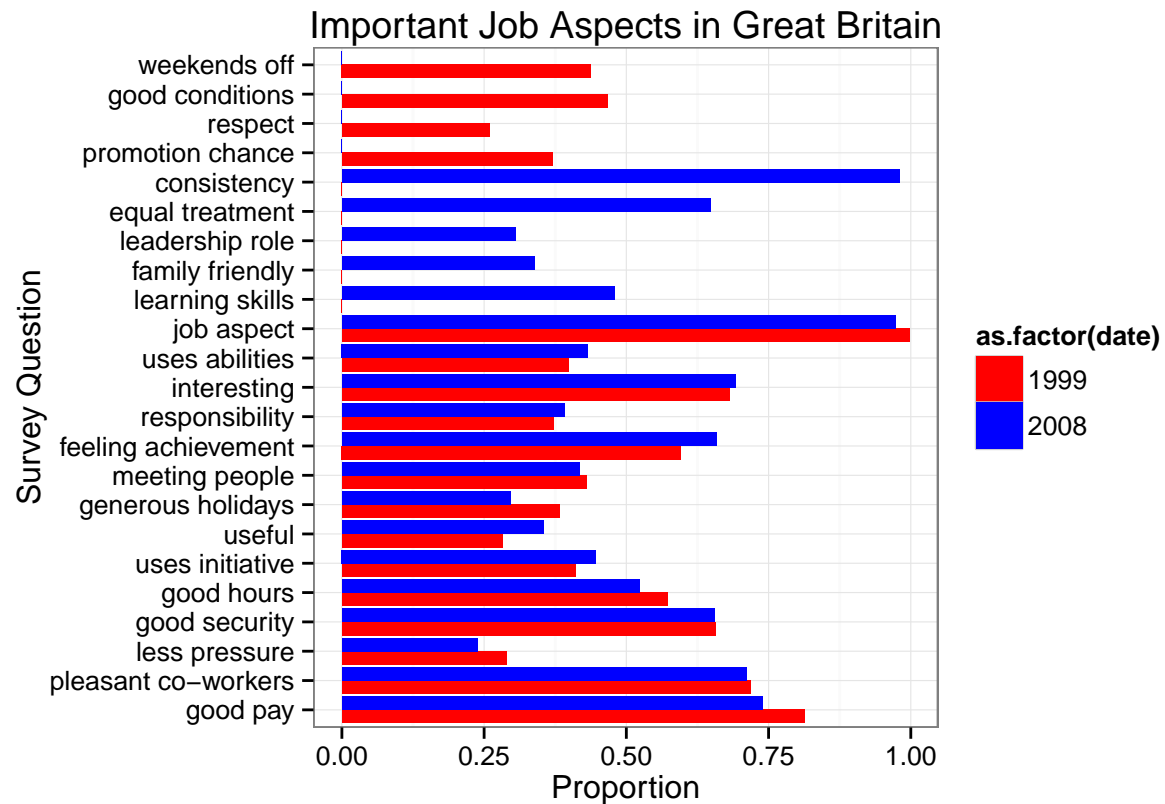


```
ggplot(agg.val, aes(variable, value)) +
    geom_point(aes(color = as.factor(date))) +
    coord_flip() +
    ylab("Proportion") +
    xlab("Survey Question") +
    ggtitle("Important Job Aspects in Great Britain") +
    scale_color_manual(values = c("red", "blue")) +
    theme_bw() #+ facet_wrap(~variable)
```

Important Job Aspects in Great Britain