Data Praticum II

Correlation between No\_HS and No\_Job is substantial at 0.489. Correlation between Chg\_Job and Chg\_Biz is not great at 0.375, however it is the biggest correlation that involves either of those two factors. Expectedly, Med\_Inc and Poverty have the greatest correlation at -0.64. This is followed by Poverty and No\_HS which has a correlation of -0.63.

Below are the correlations between DCI rank and the 7 factors plus population.

#Correlations between DCI and the rest  
res\_DCI <- cor(DCI\_2[-9], DCI\_2$DCI\_Rank)  
res\_DCI

## [,1]  
## Population -0.1230537  
## No\_Diploma 0.6604072  
## Vacancy 0.5667272  
## No\_Job 0.6697394  
**## Poverty 0.7466504  
## Med\_Inc -0.7435667**  
## Chg\_Job -0.2279098  
## Chg\_Biz -0.3546094

library(readxl)

## Warning: package 'readxl' was built under R version 3.4.3

#Importing the data set  
DCI <- read.csv("C:/Users/Daniel Tagle/Documents/Data Practicum 2/DCI Components v3.csv")  
DCI\_3 <- read.csv("C:/Users/Daniel Tagle/Documents/Data Practicum 2/DCI Components v3.csv")

#Removing superflous columns  
DCI\_3$Zip\_Code <- NULL  
DCI\_3$City <- NULL  
DCI\_3$State <- NULL  
DCI\_3$No\_Diploma\_Rank <- NULL  
DCI\_3$Vacancy\_Rank <- NULL  
DCI\_3$No\_Job\_Rank <- NULL  
DCI\_3$Poverty\_Rank <- NULL  
DCI\_3$Med\_Inc\_Rank <- NULL  
DCI\_3$Chg\_Job\_Rank <- NULL  
DCI\_3$Chg\_Biz\_Rank <- NULL  
DCI\_3$Avg\_Rank <- NULL  
DCI\_3$DCI <- NULL  
DCI\_3$DCI.246.45 <- NULL  
DCI\_3$DCI\_Sacled\_Round <- NULL  
DCI\_3$DCI\_Sacled\_Truncated <- NULL  
DCI\_3$Normalized.DCI <- NULL  
DCI\_3$DCI\_Normalized <- NULL  
  
DCI\_2 <- DCI\_3  
DCI\_2$DCI\_Rank <- DCI$DCI

#Creating the correlation matrix  
res <- cor(DCI\_3)  
round(res, 2)

## Population No\_Diploma Vacancy No\_Job Poverty Med\_Inc Chg\_Job  
## Population 1.00 0.05 -0.21 -0.09 0.06 0.13 -0.01  
## No\_Diploma 0.05 1.00 0.31 0.49 0.63 -0.53 -0.01  
## Vacancy -0.21 0.31 1.00 0.36 0.40 -0.39 -0.01  
## No\_Job -0.09 0.49 0.36 1.00 0.57 -0.46 -0.02  
## Poverty 0.06 0.63 0.40 0.57 1.00 -0.64 -0.04  
## Med\_Inc 0.13 -0.53 -0.39 -0.46 -0.64 1.00 0.07  
## Chg\_Job -0.01 -0.01 -0.01 -0.02 -0.04 0.07 1.00  
## Chg\_Biz 0.10 -0.04 -0.08 -0.07 -0.08 0.15 0.38  
## Chg\_Biz  
## Population 0.10  
## No\_Diploma -0.04  
## Vacancy -0.08  
## No\_Job -0.07  
## Poverty -0.08  
## Med\_Inc 0.15  
## Chg\_Job 0.38  
## Chg\_Biz 1.00

#Calling the Hmisc package for correlation p-values  
library(Hmisc)

## Warning: package 'Hmisc' was built under R version 3.4.4

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Warning: package 'Formula' was built under R version 3.4.1

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.4.1

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, units

#Creating object with corr and p-value  
res2 <- rcorr(as.matrix(DCI\_3))  
res2

## Population No\_Diploma Vacancy No\_Job Poverty Med\_Inc Chg\_Job  
## Population 1.00 0.05 -0.21 -0.09 0.06 0.13 -0.01  
## No\_Diploma 0.05 1.00 0.31 0.49 0.63 -0.53 -0.01  
## Vacancy -0.21 0.31 1.00 0.36 0.40 -0.39 -0.01  
## No\_Job -0.09 0.49 0.36 1.00 0.57 -0.46 -0.02  
## Poverty 0.06 0.63 0.40 0.57 1.00 -0.64 -0.04  
## Med\_Inc 0.13 -0.53 -0.39 -0.46 -0.64 1.00 0.07  
## Chg\_Job -0.01 -0.01 -0.01 -0.02 -0.04 0.07 1.00  
## Chg\_Biz 0.10 -0.04 -0.08 -0.07 -0.08 0.15 0.38  
## Chg\_Biz  
## Population 0.10  
## No\_Diploma -0.04  
## Vacancy -0.08  
## No\_Job -0.07  
## Poverty -0.08  
## Med\_Inc 0.15  
## Chg\_Job 0.38  
## Chg\_Biz 1.00  
##   
## n= 24645   
##   
##   
## P  
## Population No\_Diploma Vacancy No\_Job Poverty Med\_Inc Chg\_Job  
## Population 0.0000 0.0000 0.0000 0.0000 0.0000 0.0873   
## No\_Diploma 0.0000 0.0000 0.0000 0.0000 0.0000 0.0219   
## Vacancy 0.0000 0.0000 0.0000 0.0000 0.0000 0.0262   
## No\_Job 0.0000 0.0000 0.0000 0.0000 0.0000 0.0020   
## Poverty 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   
## Med\_Inc 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   
## Chg\_Job 0.0873 0.0219 0.0262 0.0020 0.0000 0.0000   
## Chg\_Biz 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   
## Chg\_Biz  
## Population 0.0000   
## No\_Diploma 0.0000   
## Vacancy 0.0000   
## No\_Job 0.0000   
## Poverty 0.0000   
## Med\_Inc 0.0000   
## Chg\_Job 0.0000   
## Chg\_Biz

#Function to flatten the correlation matrix  
flattenCorrMatrix <- function(cormat, pmat){  
 ut <- upper.tri(cormat)  
 data.frame(  
 row= rownames(cormat)[row(cormat)[ut]],  
 column= rownames(cormat)[col(cormat)[ut]],  
 cor= (cormat)[ut],  
 p = pmat[ut]  
 )  
}

#Flattening out the corr matrix with the p-values  
flattenCorrMatrix(res2$r, res2$P)

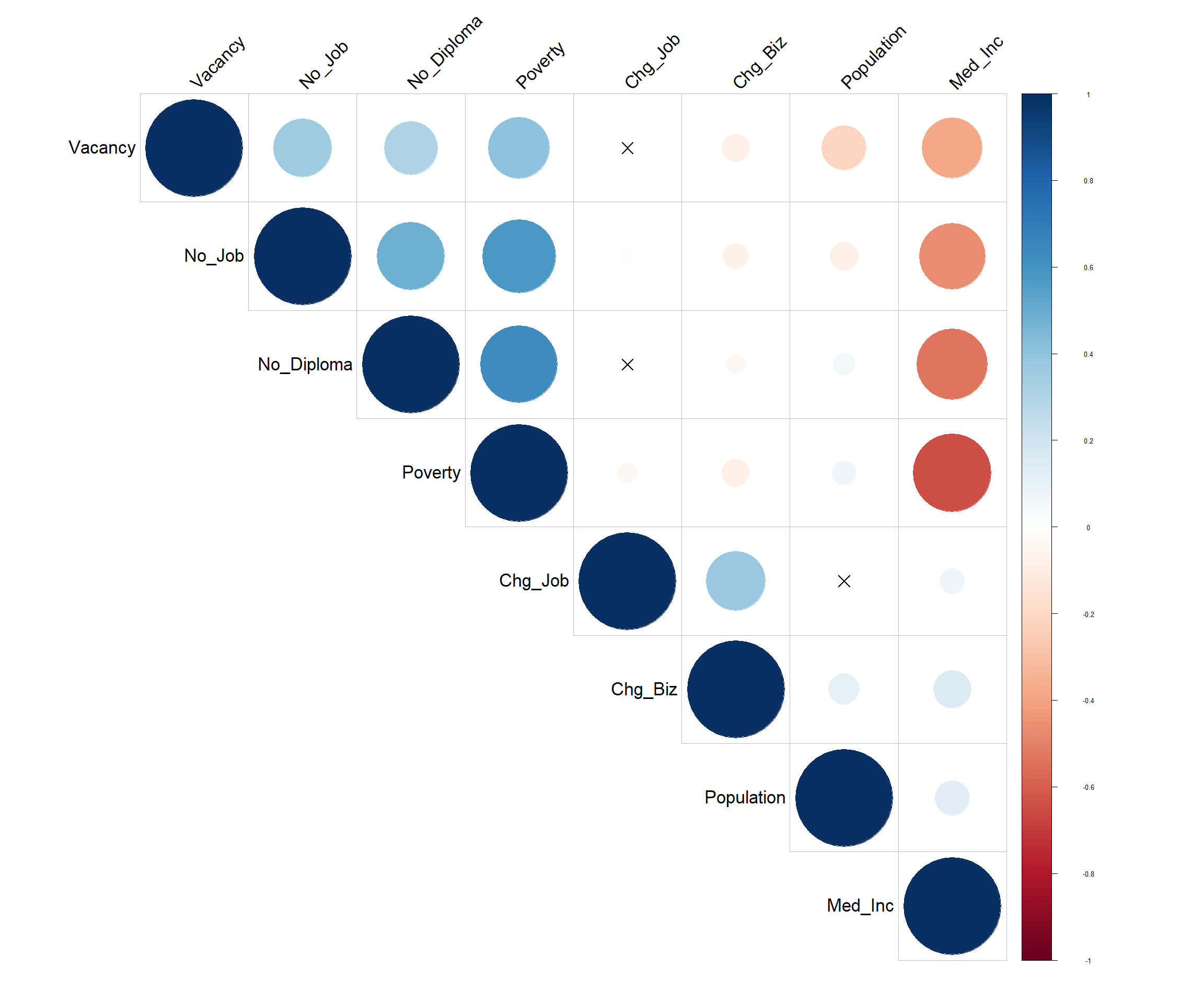
## row column cor p  
## 1 Population No\_Diploma 0.05318990 0.000000e+00  
## 2 Population Vacancy -0.21284385 0.000000e+00  
## 3 No\_Diploma Vacancy 0.30579379 0.000000e+00  
## 4 Population No\_Job -0.08810502 0.000000e+00  
## 5 No\_Diploma No\_Job 0.48850176 0.000000e+00  
## 6 Vacancy No\_Job 0.35920368 0.000000e+00  
## 7 Population Poverty 0.06370433 0.000000e+00  
## 8 No\_Diploma Poverty 0.62535563 0.000000e+00  
## 9 Vacancy Poverty 0.40217638 0.000000e+00  
## 10 No\_Job Poverty 0.57356615 0.000000e+00  
## 11 Population Med\_Inc 0.12753043 0.000000e+00  
## 12 No\_Diploma Med\_Inc -0.53374502 0.000000e+00  
## 13 Vacancy Med\_Inc -0.38527316 0.000000e+00  
## 14 No\_Job Med\_Inc -0.46158081 0.000000e+00  
## 15 Poverty Med\_Inc -0.64387527 0.000000e+00  
## 16 Population Chg\_Job -0.01089252 8.727487e-02  
## 17 No\_Diploma Chg\_Job -0.01460367 2.187126e-02  
## 18 Vacancy Chg\_Job -0.01415752 2.624628e-02  
## 19 No\_Job Chg\_Job -0.01969049 1.992825e-03  
## 20 Poverty Chg\_Job -0.04401458 4.754197e-12  
## 21 Med\_Inc Chg\_Job 0.06831566 0.000000e+00  
## 22 Population Chg\_Biz 0.10250080 0.000000e+00  
## 23 No\_Diploma Chg\_Biz -0.04098234 1.225962e-10  
## 24 Vacancy Chg\_Biz -0.08282755 0.000000e+00  
## 25 No\_Job Chg\_Biz -0.07188433 0.000000e+00  
## 26 Poverty Chg\_Biz -0.08405687 0.000000e+00  
## 27 Med\_Inc Chg\_Biz 0.15222416 0.000000e+00  
## 28 Chg\_Job Chg\_Biz 0.37522415 0.000000e+00

#Visualization (correlogram)  
library(corrplot)

## Warning: package 'corrplot' was built under R version 3.4.3

## corrplot 0.84 loaded

cplot <- corrplot(res2$r, type="upper", order="hclust", p.mat= res2$P, sig.level = 0.01, insig="pch", tl.col="black", tl.srt = 45, tl.cex = 2.1 )



#Visualization /w scatterplots  
library(PerformanceAnalytics)

## Warning: package 'PerformanceAnalytics' was built under R version 3.4.4

## Loading required package: xts

## Warning: package 'xts' was built under R version 3.4.1

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.4.1

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

##   
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':  
##   
## legend

chart.Correlation(DCI\_3, historgram=TRUE, method="pearson", pch=19)

