

```

---
title: "Lab4_RMD_Layer2"
author: "Kim Vignola"
date: "8/5/2017"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## Add libraries
```{r cars}
library(car)
library(corrplot)
```

## Load data
```{r pressure, echo=FALSE}
crime <- read.csv("crime.csv", header = TRUE, sep = ",")
```

## Transforming our dependent variable, crmrte, using a log creates a
more normal distribution for our y variable. Doing so will also increase
interpretability for our models.
```{r}
hist(log(crime$crmrte))
```

## Only 57 of 90 observatons include an indicator for geographical
region. Should we ignore these variables?
```{r}
print(nrow((subset(crime, west == 1 | central == 1 | urban == 1))))
```

## Industries such as wholesale/retail trade and the service industry
have lower median wages overall. Areas with a larger proportion of these
industries may have different profiles than other counties. The same
could be said for counties with higher wage industries (federal govt
jobs and transportation/util/communication.)
```{r}
message(sprintf("Median construction wage: %f", median(crime$wcon,
na.rm = TRUE)))
message(sprintf("Median trans, util, comm wage: %f",
median(crime$wtuc, na.rm = TRUE)))
message(sprintf("Median wholesale, retail trade wage: %f",
median(crime$wtprd, na.rm = TRUE)))
message(sprintf("Median finance, insurance, real estate wage: %f",
median(crime$wfir, na.rm = TRUE)))

```

```

message(sprintf("Median service industry wage: %f", median(crime$wser,
na.rm = TRUE)))
message(sprintf("Median manufacturing wage: %f", median(crime$wmfg,
na.rm = TRUE)))
message(sprintf("Median federal employees wage: %f",
median(crime$wfed, na.rm = TRUE)))
message(sprintf("Median state employees wage: %f", median(crime$wsta,
na.rm = TRUE)))
message(sprintf("Median local govt employees wage: %f",
median(crime$wloc, na.rm = TRUE)))
```

```

It may be interesting to group wages into total, government, business and create indicator variables for high or low. Since we are targeting local government policy, it may be interesting to look at counties with average wages above those of local government and below those of local government if we are making a recommendation on building industry (though I haven't done that here).

```

```{r}
crime$tot_wages <- (crime$wcon + crime$wtuc + crime$wtrd + crime$wfir +
crime$wser + crime$wmfg + crime$wfed + crime$wsta + crime$wloc)
crime$gov_wages <- (crime$wfed + crime$wsta + crime$wloc)
crime$bus_wages <- (crime$wcon + crime$wtuc + crime$wtrd + crime$wfir +
crime$wser + crime$wmfg)
crime$low_wages <- as.numeric(crime$tot_wages < mean(crime$tot_wages))
crime$high_wages <- as.numeric(crime$tot_wages > mean(crime$tot_wages))
```

```

Here is one correlation plot with individual categories for wages. Crime is more prevalent where taxes are high and high wage industries exist. There are possibly more arrests when there are more police per capita.

```

```{r}
corrplot(cor(crime[, (names(crime) %in% c("crmrte", "density",
"prbarr", "prbconv", "prbpris", "avgsgen", "polpc", "taxpc", "pctmin80",
"mix", "pctymle", "wcon", "wtuc", "wtrd", "wfir", "wser", "wmfg",
"wfed", "wsta", "wloc"))]), method="number")
```

```

The following correlation plot aggregates wages. Crime rates are higher where government wages are higher (though how we can recommend policy around that may require some innovative thinking).

```

```{r}
corrplot(cor(crime[, (names(crime) %in% c("crmrte", "density",
"prbarr", "prbconv", "prbpris", "avgsgen", "polpc", "taxpc", "pctmin80",
"mix", "pctymle", "tot_wages", "gov_wages", "bus_wages", "low_wages",
"high_wages"))]), method="number")
```

```

Questions:

Is there some way to transform the prison arrests, convictions, sentencing metrics to create one meaningful one? I haven't found one but am not finding much meaningful in these metrics, apart from the fact that arrests are higher in high wage areas.

Some of the variable descriptions are unclear. I sent Ryan a note on Slack regarding the "mix" variable. But we should gather our questions (agree with likely error in Days vs. Years in prison sentence) and send those to Ryan.