

# CKME136 W17 Capstone Project

Leung Ming Kevin Tam ID 500209198

March 19, 2017

```
setwd("D:/OneDrive/Dropbox/CKME136")
tax06.data<-read.csv(file="property_tax_report_csv2006.csv",header=T,sep=",")
tax07.data<-read.csv(file="property_tax_report_csv2007.csv",header=T,sep=",")
tax08.data<-read.csv(file="property_tax_report_csv2008.csv",header=T,sep=",")
tax09.data<-read.csv(file="property_tax_report_csv2009.csv",header=T,sep=",")
tax10.data<-read.csv(file="property_tax_report_csv2010.csv",header=T,sep=",")
tax11.data<-read.csv(file="property_tax_report_csv2011.csv",header=T,sep=",")
tax12.data<-read.csv(file="property_tax_report_csv2012.csv",header=T,sep=",")
tax13.data<-read.csv(file="property_tax_report_csv2013.csv",header=T,sep=",")
tax14.data<-read.csv(file="property_tax_report_csv2014.csv",header=T,sep=",")
tax15.data<-read.csv(file="property_tax_report_csv2015.csv",header=T,sep=",")
tax16.data<-read.csv(file="property_tax_report_csv2016.csv",header=T,sep=",")
tax17.data<-read.csv(file="property_tax_report.csv",header=T,sep=",")
```

check for na fields

```
tax06.data[is.na(tax06.data$TAX_ASSESSMENT_YEAR),]
tax07.data[is.na(tax07.data$TAX_ASSESSMENT_YEAR),]
tax08.data[is.na(tax08.data$TAX_ASSESSMENT_YEAR),]
tax09.data[is.na(tax09.data$TAX_ASSESSMENT_YEAR),]
tax10.data[is.na(tax10.data$TAX_ASSESSMENT_YEAR),]
tax11.data[is.na(tax11.data$TAX_ASSESSMENT_YEAR),]
tax12.data[is.na(tax12.data$TAX_ASSESSMENT_YEAR),]
tax13.data[is.na(tax13.data$TAX_ASSESSMENT_YEAR),]
tax14.data[is.na(tax14.data$TAX_ASSESSMENT_YEAR),]
tax15.data[is.na(tax15.data$TAX_ASSESSMENT_YEAR),]
tax16.data[is.na(tax16.data$TAX_ASSESSMENT_YEAR),]
tax17.data[is.na(tax17.data$TAX_ASSESSMENT_YEAR),]
```

confirm that 5566 records for tax14.data, 5478 records for tax15.data, 3734 records for tax16.data, 1024 records for tax17.data are missing values of TAX\_ASSESSMENT\_YEAR column

```
sum(is.na(tax14.data$TAX_ASSESSMENT_YEAR))
```

```
## [1] 5566
```

```
sum(is.na(tax15.data$TAX_ASSESSMENT_YEAR))
```

```
## [1] 5478
```

```
sum(is.na(tax16.data$TAX_ASSESSMENT_YEAR))
```

```
## [1] 3734
```

```
sum(is.na(tax17.data$TAX_ASSESSMENT_YEAR))
```

```
## [1] 1024
```

replacing missing values for TAX\_ASSESSMENT\_YEAR field

```
tax14.data$TAX_ASSESSMENT_YEAR[is.na(tax14.data$TAX_ASSESSMENT_YEAR)] <- 2014
tax15.data$TAX_ASSESSMENT_YEAR[is.na(tax15.data$TAX_ASSESSMENT_YEAR)] <- 2015
```

```
tax16.data$TAX_ASSESSMENT_YEAR[is.na(tax16.data$TAX_ASSESSMENT_YEAR)] <- 2016  
tax17.data$TAX_ASSESSMENT_YEAR[is.na(tax17.data$TAX_ASSESSMENT_YEAR)] <- 2017
```

binding all tax datasets

```
tax.data <- rbind(tax06.data,tax07.data)  
tax.data <- rbind(tax.data,tax08.data)  
tax.data <- rbind(tax.data,tax09.data)  
tax.data <- rbind(tax.data,tax10.data)  
tax.data <- rbind(tax.data,tax11.data)  
tax.data <- rbind(tax.data,tax12.data)  
tax.data <- rbind(tax.data,tax13.data)
```

add back missing columns for tax06.data to tax13.data

```
tax.data$ZONE_NAME <- ""  
tax.data$ZONE_CATEGORY <- ""  
tax.data$PREVIOUS_LAND_VALUE <- ""  
tax.data$PREVIOUS_IMPROVEMENT_VALUE <- ""  
  
tax.data <- rbind(tax.data,tax14.data)  
tax.data <- rbind(tax.data,tax15.data)  
tax.data <- rbind(tax.data,tax16.data)  
tax.data <- rbind(tax.data,tax17.data)
```

remove records with empty CURRENT\_LAND\_VALUE (first check how many rows has empty CURRENT\_LAND\_VALUE)

```
sum(is.na(tax.data$CURRENT_LAND_VALUE))
```

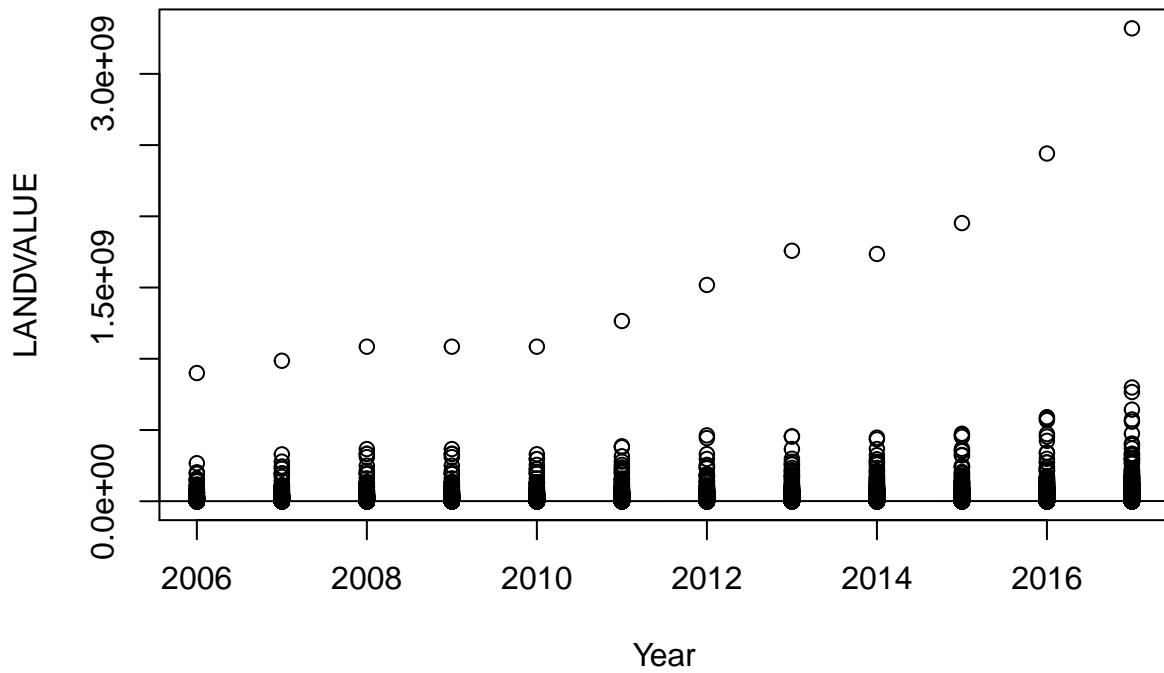
```
## [1] 15802
```

```
tax.data1 <- tax.data[!(is.na(tax.data$CURRENT_LAND_VALUE)),]
```

```
x <- as.vector(tax.data1$TAX_ASSESSMENT_YEAR)  
y <- as.vector(tax.data1$CURRENT_LAND_VALUE)
```

create formula with Linear Regression for the prediction to CURRENT\_LAND\_VALUE and plot

```
lm.price = lm(formula = y ~ x, data = tax.data1)  
plot(x,y, xlab="Year", ylab="LANDVALUE")  
abline(lm.price)
```



Predicated CURRENT\_LAND\_VALUE for 2018 is

```
-161818411 + (2018 * 80907)
```

```
## [1] 1451915
```

Predicated CURRENT\_LAND\_VALUE for 2019 is

```
-161818411 + (2019 * 80907)
```

```
## [1] 1532822
```

Predicated CURRENT\_LAND\_VALUE for 2020 is

```
-161818411 + (2020 * 80907)
```

```
## [1] 1613729
```