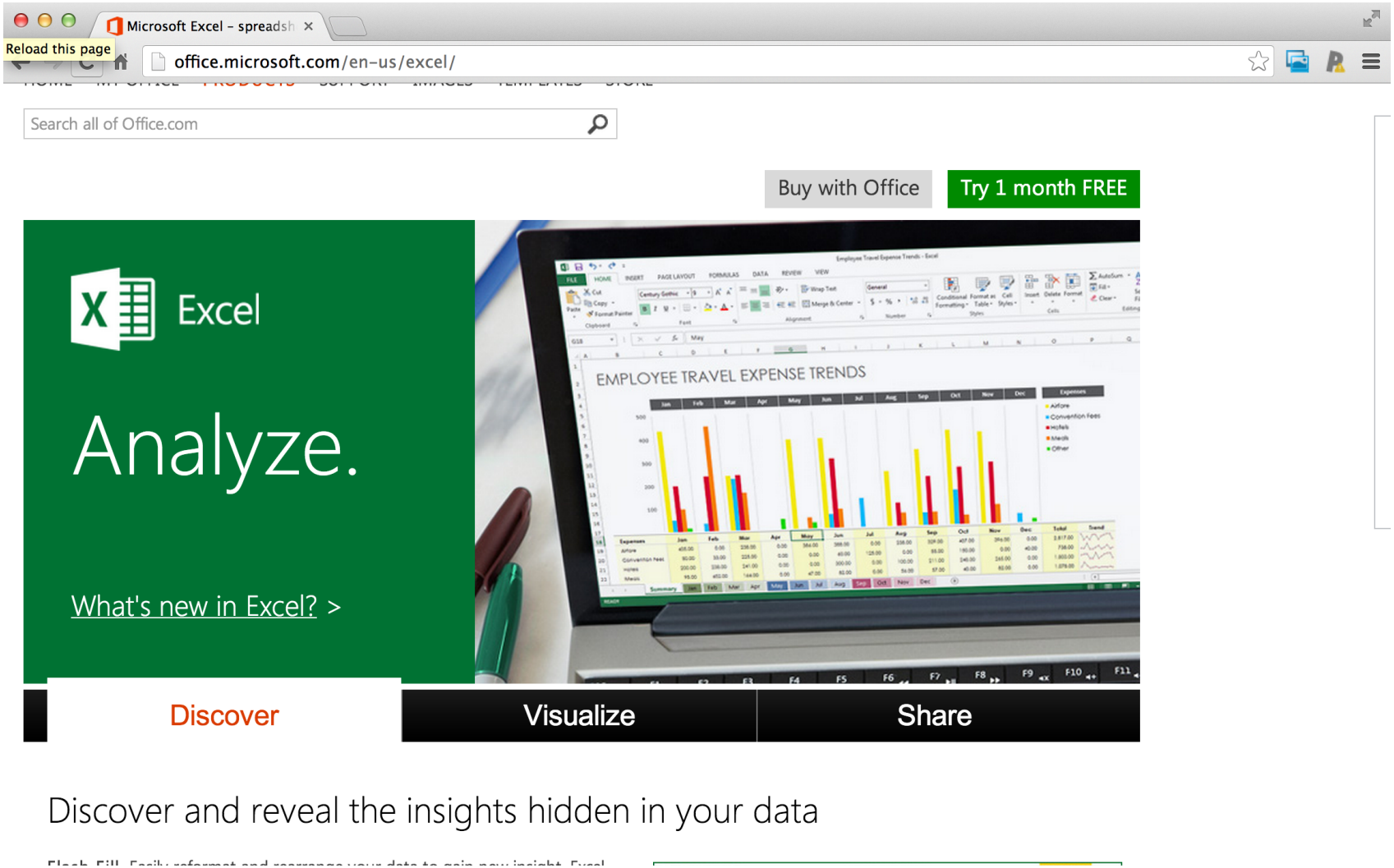
# Reading Excel files

Jeffrey Leek

May 17, 2016

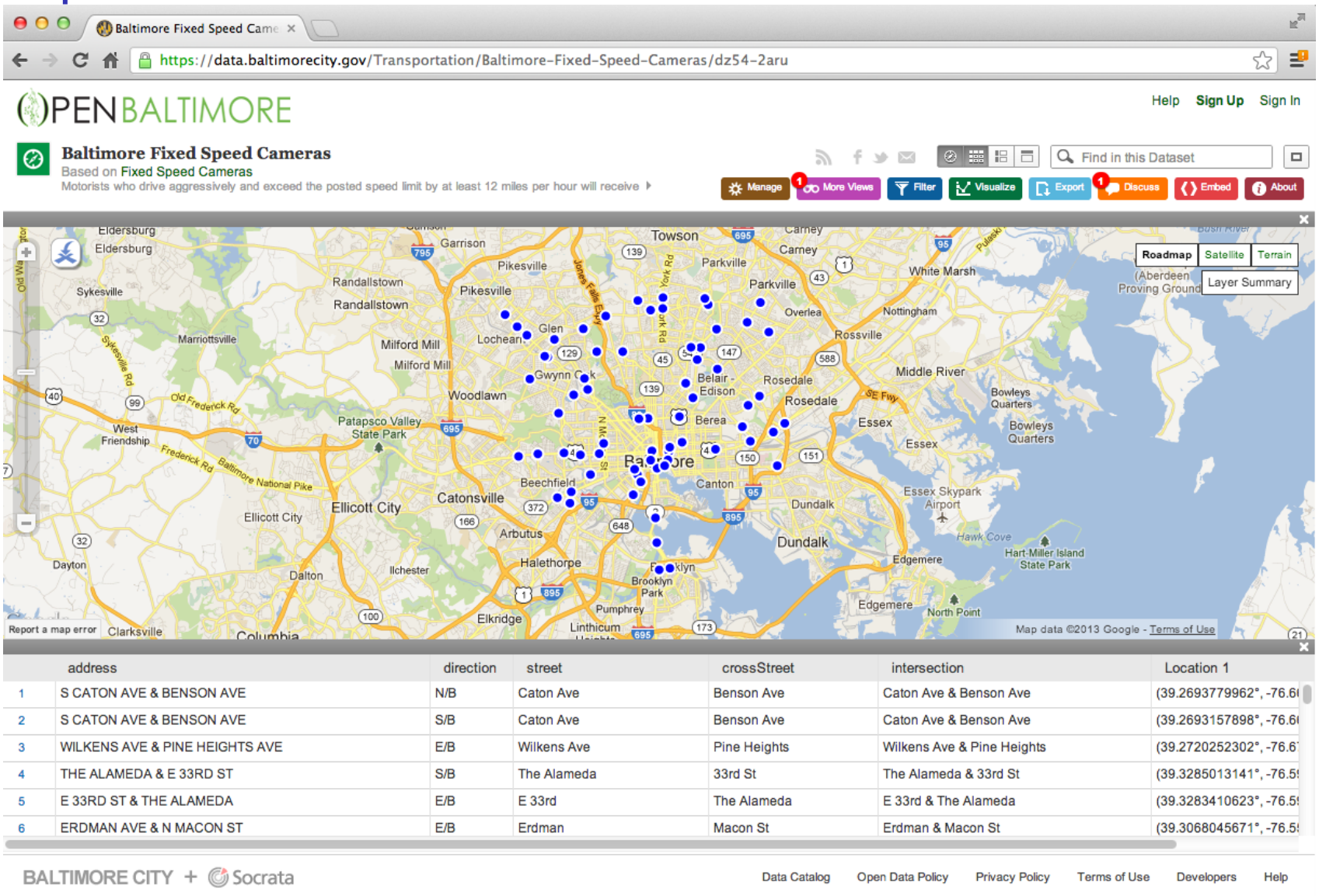
# Excel files

*Still probably the most widely used format for sharing data*




`http://office.microsoft.com/en-us/excel/`

# Example - Baltimore camera data




https://data.baltimorecity.gov/Transportation/Baltimore-Fixed-Speed-Cameras/dz54-2aru

# Download the file to load

```
if(!file.exists("data")){dir.create("data")}
fileUrl <- "https://data.baltimorecity.gov/api/views/dz54-2
download.file(fileUrl,destfile="./data/cameras.xlsx",metho
dateDownloaded <- date()
```

# read.xlsx(), read.xlsx2() {xlsx package}

```
library(xlsx)
```

```
## Loading required package: rJava

## Loading required package: methods

## Loading required package: xlsxjars
```

```
cameraData <- read.xlsx("./data/cameras.xlsx",sheetIndex=1
head(cameraData)
```

```
##                                address direction      street
## 1        S CATON AVE & BENSON AVE       N/B   Caton Ave
## 2        S CATON AVE & BENSON AVE       S/B   Caton Ave
## 3 WILKENS AVE & PINE HEIGHTS AVE       E/B Wilkens Ave H
## 4         THE ALAMEDA & E 33RD ST       S/B The Alameda
## 5         E 33RD ST & THE ALAMEDA       E/B      E 33rd
## 6         ERDMAN AVE & N MACON ST       E/B      Erdman
```

# Reading specific rows and columns

```
colIndex <- 2:3
rowIndex <- 1:4
cameraDataSubset <- read.xlsx("./data/cameras.xlsx",sheetIn
                              colIndex=colIndex,rowIndex=ro
cameraDataSubset
```

```
##   direction      street
## 1       N/B   Caton Ave
## 2       S/B   Caton Ave
## 3       E/B Wilkens Ave
```

# Further notes

- The *write.xlsx* function will write out an Excel file with similar arguments.
- *read.xlsx2* is much faster than *read.xlsx* but for reading subsets of rows may be slightly unstable.
- The XLConnect package has more options for writing and manipulating Excel files
- The XLConnect vignette is a good place to start for that package
- In general it is advised to store your data in either a database or in comma separated files (.csv) or tab separated files (.tab/.txt) as they are easier to distribute.