

# Reading Excel files

Jeffrey Leek

May 17, 2016

# Excel files

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

The image shows a browser window with the URL [office.microsoft.com/en-us/excel/](http://office.microsoft.com/en-us/excel/). Below the browser window is a green banner with the Excel logo and the text "Analyze." and "What's new in Excel? >". To the right of the banner is a laptop displaying an Excel spreadsheet titled "EMPLOYEE TRAVEL EXPENSE TRENDS". The spreadsheet shows a bar chart and a data table. Below the banner and laptop are three buttons: "Discover", "Visualize", and "Share".

Buy with Office Try 1 month FREE

Excel

Analyze.

What's new in Excel? >

Discover Visualize Share

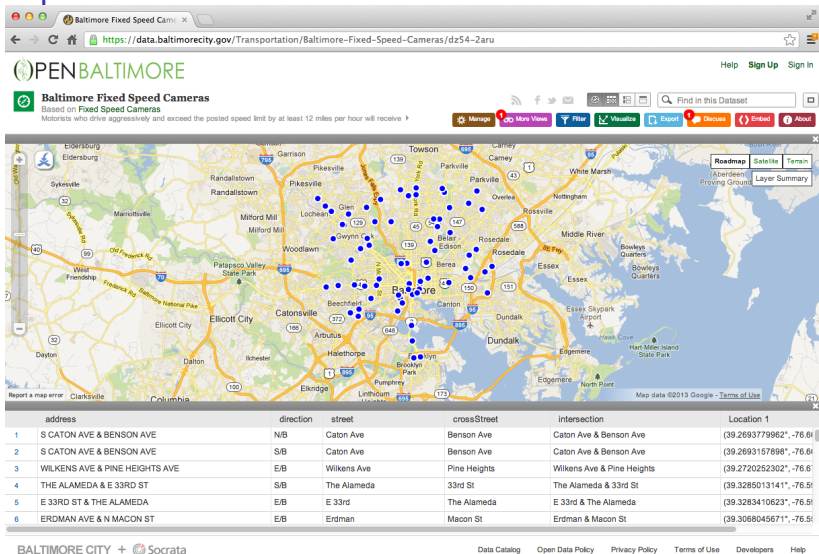
EMPLOYEE TRAVEL EXPENSE TRENDS

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Expenses	1000	1200	1500	1800	2000	2200	2500	2800	3000	3200	3500	3800	30000
Travel	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	12000
Meals	300	350	400	450	500	550	600	650	700	750	800	850	6000
Hotels	200	250	300	350	400	450	500	550	600	650	700	750	5000
Transportation	100	120	140	160	180	200	220	240	260	280	300	320	2000

Discover and reveal the insights hidden in your 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",method  
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	Wilkins Ave
## 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",sheetIndex=1,  
                               colIndex=colIndex,rowIndex=rowIndex)  
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.