



# Downloading files

Jeffrey Leek

Johns Hopkins Bloomberg School of Public Health

# Get/set your working directory

- A basic component of working with data is knowing your working directory
- The two main commands are `getwd()` and `setwd()`.
- Be aware of relative versus absolute paths
  - Relative - `setwd("data")`, `setwd("../")`
  - Absolute- `setwd("/Users/jtleek/data/")`
- Important difference in Windows `setwd("c:\\users\\Andrew\\Downloads")`

# Checking for and creating directories

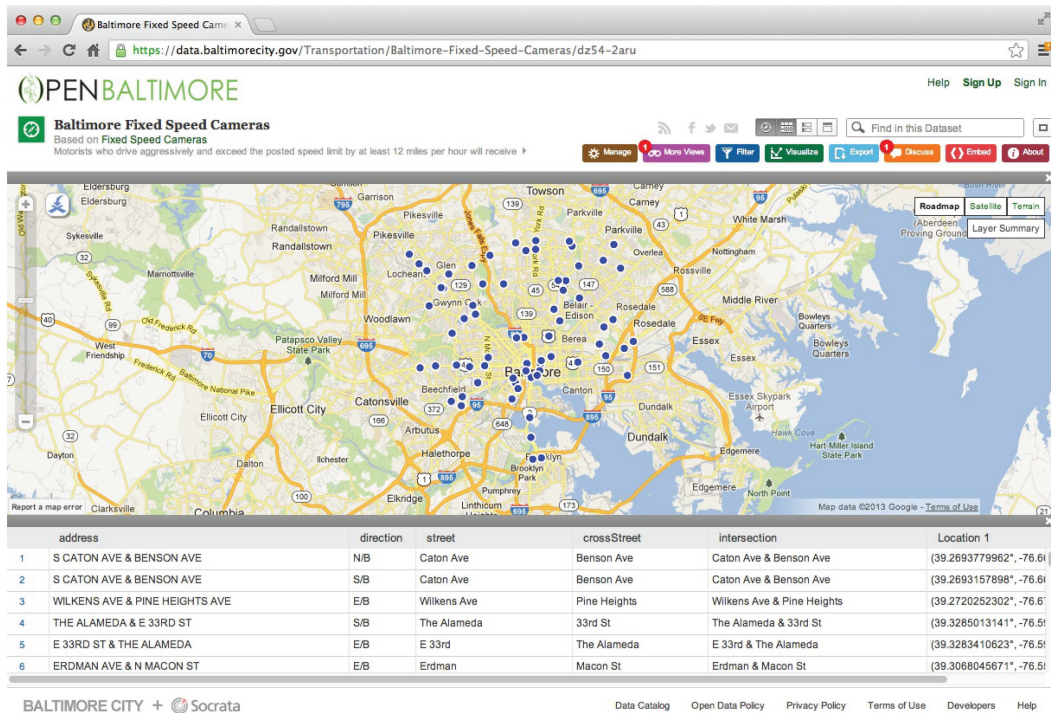
- `file.exists("directoryName")` will check to see if the directory exists
- `dir.create("directoryName")` will create a directory if it doesn't exist
- Here is an example checking for a "data" directory and creating it if it doesn't exist

```
if (!file.exists("data")) {  
  dir.create("data")  
}
```

# Getting data from the internet - download. fileO

- Downloads a file from the internet
- Even if you could do this by hand, helps with reproducibility
- Important parameters are *url*, *destfile*, *method*
- Useful for downloading tab-delimited, csv, and other files

# Example - Baltimore camera data



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

# Example - Baltimore camera data

**OPENBALTIMORE**  
Baltimore Fixed Speed Cameras  
Based on Fixed Speed Cameras  
Motorists who drive aggressively and exceed the posted speed limit by at least 12 miles per hour will receive

Map data ©2015 Google - Terms of Use

	address	direction	street	crossStreet	Intersection
1	S CATON AVE & BENSON AVE	N/B	Caton Ave	Benson Ave	Caton Ave & Be
2	S CATON AVE & BENSON AVE	S/B	Caton Ave	Benson Ave	Caton Ave & Be
3	WILKENS AVE & PINE HEIGHTS AVE	E/B	Wilkins Ave	Pine Heights	Wilkins Ave & i
4	THE ALAMEDA & E 33RD ST	S/B	The Alameda	33rd St	The Alameda &
5	E 33RD ST & THE ALAMEDA	E/B	E 33rd	The Alameda	E 33rd & The Al
6	ERDMAN AVE & N MACON ST	E/B	Erdman	Macon St	Erdman & Maco

BALTIMORE CITY + Socrata

Data Catalog Open Data Policy Privacy Policy Terms of Use Developers Help

Export  
API  
Download  
Download a copy of this dataset in a static format  
Download As  
Open Link in New Tab  
Open Link in New Window  
Open Link in Incognito Window  
Save Link As...  
Copy Link Address  
Copy  
Search Google for 'CSV'  
Inspect Element  
Look Up in Dictionary  
Speech  
Search With Google  
Add to iTunes as a Spoken Track

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

# Download a file from the web

```
fileUrl <- "https://data.baltimorecity.gov/api/views/dz54-2aru/rows.csv?accessType=DOWNLOAD"  
download.file(fileUrl, destfile = "./data/cameras.csv", method = "curl")  
list.files("./data")
```

```
## [1] "cameras.csv"
```

```
dateDownloaded <- date()  
dateDownloaded
```

```
## [1] "Sun Jan 12 21:37:44 2014"
```

# Some notes about `download.file()`

- If the url starts with *http* you can use `download.file()`
- If the url starts with *https* on Windows you may be ok
- If the url starts with *https* on Mac you may need to set *method="curl"*
- If the file is big, this might take a while
- Be sure to record when you downloaded.