

Aquisição de Arquivos e Datasets

Downloading files

Delermundo Branquinho Filho

Obter / definir seu diretório de trabalho

- Um componente básico de trabalhar com dados é conhecer o seu diretório de trabalho
 - Os dois comandos principais são `'getwd ()'` e `'setwd ()'`.
 - Esteja ciente dos caminhos relativos versus absolutos
 - **Relative** - `'setwd (" ./ data ")'`, `'setwd (" ../ ")'`
 - **Absolute** - `'setwd (" / Users / jtleek / data / ")'`
 - Diferença importante no Windows `'setwd (" C: \\ Users \\ Andrew \\ Downloads ")'`
-

Verificando e criando diretórios

- `'File.exists (" directoryName ")'` verificará se o diretório existe
- `'Dir.create (" directoryName ")'` criará um diretório se ele não existir
- Aqui está um exemplo de verificação de um diretório “data” e criá-lo se ele não existir

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

Obtendo dados da Internet - `download.file()`

- Descarrega um ficheiro da Internet
 - Mesmo se você pudesse fazer isso à mão, ajuda com a reprodutibilidade
 - Parâmetros importantes são *url*, *destfile*, *method*
 - Útil para download de tabulações, csv e outros arquivos
-

Download de um arquivo da 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] "Mon Mar 27 10:43:19 2017"
```

Algumas observações sobre `download.file()`

- Se o URL começar com *http* você pode usar `download.file()`
- Se a url começa com *https* no Windows você pode estar ok
- Se a url começa com *https* no Mac você pode precisar definir *method = "curl"*
- Se o arquivo for grande, isso pode levar algum tempo
- Certifique-se de gravar quando você baixar.