

# Aquisição de Arquivos e Datasets

Reading local flat files

*Delermendo Branquinho Filho*

Faça o download do arquivo para carregar

```
if(!dir.exists("data")) {  
  dir.create("data")  
}  
fileUrl <- "https://data.baltimorecity.gov/api/views/dz54-2aru/rows.csv?accessType=DOWNLOAD"  
download.file(fileUrl,destfile="date/cameras.csv",method="curl")  
dateDownloaded <- date()  
dateDownloaded  
  
## [1] "Mon Mar 27 10:42:30 2017"  
file.info("date/cameras.csv")  
  
##              size isdir mode mtime ctime atime  exe  
## date/cameras.csv   NA     NA <NA>  <NA>  <NA>  <NA>  <NA>
```

---

Carregando arquivos texto - read.table ()

- Esta é a função principal para a leitura de dados em R
  - Flexível e robusto, mas requer mais parâmetros
  - Lê os dados na RAM - grandes dados podem causar problemas
  - Parâmetros importantes *file*, *header*, *sep*, *row.names*, *nrows*
  - Relacionado: *read.csv ()* , *read.csv2 ()*
- 

Baltimore exemplo

```
cameraData <- read.table("./data/cameras.csv",sep="," ,header=TRUE)  
head(cameraData)  
  
##              address direction      street crossStreet  
## 1      S CATON AVE & BENSON AVE      N/B   Caton Ave   Benson Ave  
## 2      S CATON AVE & BENSON AVE      S/B   Caton Ave   Benson Ave  
## 3 WILKENS AVE & PINE HEIGHTS AVE      E/B Wilkens Ave Pine Heights  
## 4      THE ALAMEDA & E 33RD ST      S/B The Alameda      33rd St  
## 5      E 33RD ST & THE ALAMEDA      E/B      E 33rd   The Alameda  
## 6      ERDMAN AVE & N MACON ST      E/B      Erdman     Macon St  
##              intersection      Location.1  
## 1      Caton Ave & Benson Ave (39.2693779962, -76.6688185297)  
## 2      Caton Ave & Benson Ave (39.2693157898, -76.6689698176)  
## 3 Wilkens Ave & Pine Heights (39.2720252302, -76.676960806)  
## 4      The Alameda & 33rd St (39.3285013141, -76.5953545714)  
## 5      E 33rd & The Alameda (39.3283410623, -76.5953594625)
```

## Alguns parâmetros mais importantes

- *quote* - você pode dizer a R se existem valores citados *quote* = “” significa sem aspas.
- *na.strings* - define o caractere que representa um valor ausente.
- *nrows* - quantas linhas a ler do arquivo (por exemplo, *nrows* = 10 lê 10 linhas).
- *skip* - número de linhas a saltar antes de começar a ler

*Na minha experiência, o maior problema com a leitura de arquivos planos são aspas “ou” colocados em valores de dados, definindo citações = “” muitas vezes resolve these.*