

Symbols and functions from today's session

R

Import data (CSV)

```
df <- read.csv(...) # imports the data at '...' as a data frame
```

For example,

```
df <- read.csv('https://tinyurl.com/2s3p9s2z') #loads the CSV from this link
another_df <- read.csv('senderdata.csv') #loads the same data, but from my computer
```

Inspect data

```
head(df) #shows first lines of the data frame 'df'
```

```
##   pairnum      IP trial display  label
## 1      1 67.85.42.18      1      0    zop
## 2      1 67.85.42.18      2      3 zopudon
## 3      1 67.85.42.18      3      0    zop
## 4      1 67.85.42.18      4      0 zopekil
## 5      1 67.85.42.18      5      2 zopudon
## 6      1 67.85.42.18      6      1 zopekil
```

```
tail(df) #shows the last lines of df
```

```
##      pairnum      IP trial display  label
## 1275      20 74.87.33.130      27      1    zop
## 1276      20 74.87.33.130      28      2 zopudon
## 1277      20 74.87.33.130      29      0    zop
## 1278      20 74.87.33.130      30      0    zop
## 1279      20 74.87.33.130      31      1    zop
## 1280      20 74.87.33.130      32      2 zopudon
```

```
nrow(df) #number of rows of df
```

```
## [1] 1280
```

```
colnames(df) #column names of df
```

```
## [1] "pairnum" "IP"      "trial"  "display" "label"
```

Extract and manipulate column data

```
data <- df$column_name #the operator $ extracts the column named "column_name" from "df"
```

For example,

```
labels_used <- df$label #data in the "label" column
```

```
unique(labels_used) #unique values in "labels_used"
```

```
## [1] zop      zopudon zopekil  
## Levels: zop zopekil zopudon
```

Filtering

```
library(dplyr) #library with functions to filter/select data
```

```
filter(x,y == z) #takes the data frame x and selects all lines in which column y has value z
```

For example,

```
df_zopudon <- filter(df, label == 'zopudon')  
head(df_zopudon)
```

```
##   pairnum      IP trial display  label  
## 1      1 67.85.42.18      2      3 zopudon  
## 2      1 67.85.42.18      5      2 zopudon  
## 3      1 67.85.42.18     11      2 zopudon  
## 4      1 67.85.42.18     13      3 zopudon  
## 5      1 67.85.42.18     14      3 zopudon  
## 6      1 67.85.42.18     18      2 zopudon
```

```
nrow(df_zopudon)
```

```
## [1] 369
```