

Wrangling Data in R
Week 11 – Thursday
Applied Regression Analysis (STAT 757)

Paul J. Hurtado

31 Mar, 2016

Importing Data in R

Base R functions: `read.csv`, `read.table`.

Packages: `reshape2`, `tidyr`, `dplyr`, `zoo`, `lubridate`...

Resources:

- ① Data Input at Quick-R: <http://www.statmethods.net/input/>
- ② R manual: **R Data Import/Export**.
<https://cran.r-project.org/manuals.html>
- ③ New readxl package by Hadley Wickham
- ④ RStudio cheat sheet:
<http://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>
- ⑤ See also this comparison
<http://rpubs.com/paul4forest/reshape2tidyrdplyr>
- ⑥ For more about tidyr
<http://blog.rstudio.org/2014/07/22/introducing-tidyr/>

Data Frames

```
mydat = data.frame(A = 3:1, B = c("a", "b", "c"), C = runif(3))
str(mydat)
```

```
## 'data.frame': 3 obs. of 3 variables:
## $ A: int 3 2 1
## $ B: Factor w/ 3 levels "a","b","c": 1 2 3
## $ C: num 0.32 0.287 0.527
```

```
mydat[2, 2] # row, column addressing
```

```
## [1] b
## Levels: a b c
```

```
mydat[, 2] # all rows, 2nd column
```

```
## [1] a b c
## Levels: a b c
```

```
mydat[c(1, 3), ] # 1st and 3rd rows, all columns
```

```
##   A B      C
## 1 3 a 0.3197704
## 3 1 c 0.5274163
```

Data Frames (cont'd)

```
names(mydat)

## [1] "A" "B" "C"

mydat$B # access columns of data via column names

## [1] a b c
## Levels: a b c

mydat[, c(TRUE, FALSE, TRUE)] # Subset columns with logical vectors

##      A          C
## 1 3 0.3197704
## 2 2 0.2872676
## 3 1 0.5274163

mydat[mydat$A >= 2, ] # useful for subsetting data!

##      A B          C
## 1 3 a 0.3197704
## 2 2 b 0.2872676
```

Exercise

Go to www.ebird.org > **Explore Data** > **Barcharts** and download three summary barchart data sets:

Nevada for 2008, 2010, 2012

Open using `read.csv()` and add a **Year** column to each data frame.

Optional: Also extract the sample size information into a separate data frame.

Merge these into a single data frame.

Convert to long format, collapsing the different **Week** columns into a single column.