Introduction to R for Data Management and Analysis

Marcel Ramos, MPH

Tuesday, June 14, 2016

Notes on last Thursday's lecture

- Examples with pipes
- Reshaping your data
- Merge alternatives

```
library(nycflights13); library(dplyr)
flights %>% group_by(carrier) %>%
  summarise(avg_depdelay = mean(dep_delay, na.rm = TRUE),
            count = n()) %>% left_join(airlines) %>%
    arrange(avg_depdelay) %>% head

## Source: local data frame [6 x 4]
##
##    carrier avg_depdelay count                    name
##      (chr)         (dbl) (int)                   (chr)
## 1       US     3.782418 20536          US Airways Inc.
## 2       HA     4.900585   342     Hawaiian Airlines Inc.
## 3       AS     5.804775   714      Alaska Airlines Inc.
## 4       AA     8.586016 32729    American Airlines Inc.
## 5       DL     9.264505 48110        Delta Air Lines Inc.
## 6       MQ    10.552041 26397               Envoy Air
```

Reshaping data using `gather`

```r
data(iris); library(tidyr)
longdata <- gather(tbl_df(iris), key = measure, n,
  Sepal.Length:Petal.Width) %>% separate(measure, c("type",
    "dimension"))
longdata %>% group_by(Species, type, dimension) %>%
  summarise(avg_dim = mean(n, na.rm = TRUE))
```

```
## Source: local data frame [12 x 4]
## Groups: Species, type [?]
##
##        Species  type dimension avg_dim
##         (fctr) (chr)     (chr)   (dbl)
## 1       setosa Petal    Length   1.462
## 2       setosa Petal     Width   0.246
## 3       setosa Sepal    Length   5.006
## 4       setosa Sepal     Width   3.428
```

Pew example

```
library(readr)
# (pew <- read_csv("https://raw.githubusercontent.com/LiNk-NY/
```