Heatmaps - the gene expression edition

Jeff Oliver

13 September, 2017

[INTRODUCTORY SENTENCE]

Learning objectives

1. one
2. two
3. three

[DESCRIPTION OR MOTIVATION; 2-4 sentences that would be used for an announcement]

Getting started

[TOPIC ONE]

[TOPIC TWO]

Install two packages (in console)

```r
install.packages("ggplot2")
install.packages("tidyr")
```

Create places for data and output

```r
dir.create("data")
dir.create("output")
```

Download data and put it in the `data` folder. Data is at: http://tinyurl.com/flu-expression-data or https://jcoliver.github.io/learn-r/data/GSE68849-expression.csv

Create a script that will save all our work; also, load additional packages

```r
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")
```

Read in the data

```r
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)
```

Investigate data with `summary`

```r
summary(exp.data)
```

```
##    subject            treatment             IFNA5
##  Length:10          Length:10           Min.   :   78.12
##  Class :character   Class :character    1st Qu.:   86.80
##  Mode  :character   Mode  :character    Median :  3578.65
##                                         Mean   :  5019.12
##                                         3rd Qu.:  9093.19
##                                         Max.   : 15018.41
##      IFNA13             IFNA2              SPIN1            ZNF451
##  Min.   :   98.34   Min.   :   100.7   Min.   :104.2   Min.   :232.2
##  1st Qu.:  110.58   1st Qu.:   124.9   1st Qu.:109.9   1st Qu.:289.3
##  Median :  7765.56   Median :  9665.8   Median :119.4   Median :302.4
##  Mean   :  8617.42   Mean   : 10647.0   Mean   :118.2   Mean   :333.9
##  3rd Qu.: 16655.33   3rd Qu.: 20046.6   3rd Qu.:123.3   3rd Qu.:325.3
##  Max.   : 18974.16   Max.   : 24029.1   Max.   :134.8   Max.   :569.3
##      IFNA16             RASSF1           IFNW1              MSR1
##  Min.   :   101.2   Min.   :244.8   Min.   :   89.26   Min.   : 89.58
##  1st Qu.:   158.9   1st Qu.:291.6   1st Qu.:   95.79   1st Qu.: 95.48
##  Median :  8968.3   Median :330.5   Median : 2883.71   Median :100.83
##  Mean   : 10378.3   Mean   :323.5   Mean   : 3834.03   Mean   :103.43
##  3rd Qu.: 20543.8   3rd Qu.:353.3   3rd Qu.: 7297.51   3rd Qu.:108.38
##  Max.   : 23060.5   Max.   :386.8   Max.   : 9198.55   Max.   :125.94
##     MIR1976
##  Min.   : 82.85
##  1st Qu.: 87.79
##  Median :103.58
##  Mean   :100.16
##  3rd Qu.:108.23
##  Max.   :120.63
```

Transform data to long format

```r
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)
```

```
exp.long <- gather(data = exp.data,
                   key = gene,
                   value = expression,
                   -subject, -treatment)
```

Look at the top of the long data:

```
head(exp.long)
```

```
##      subject treatment  gene  expression
## 1 GSM1684095   control IFNA5    83.12937
## 2 GSM1684096 influenza IFNA5 10096.47000
## 3 GSM1684097   control IFNA5    97.80374
## 4 GSM1684098 influenza IFNA5  8180.98900
## 5 GSM1684099   control IFNA5    81.70878
## 6 GSM1684100 influenza IFNA5  7054.90200
```

Plot!

```
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)

exp.long <- gather(data = exp.data,
                   key = gene,
                   value = expression,
                   -subject, -treatment)

exp.heatmap <- ggplot(data = exp.long, mapping = aes(x = subject,
                                                     y = gene,
                                                     fill = expression))
exp.heatmap
```
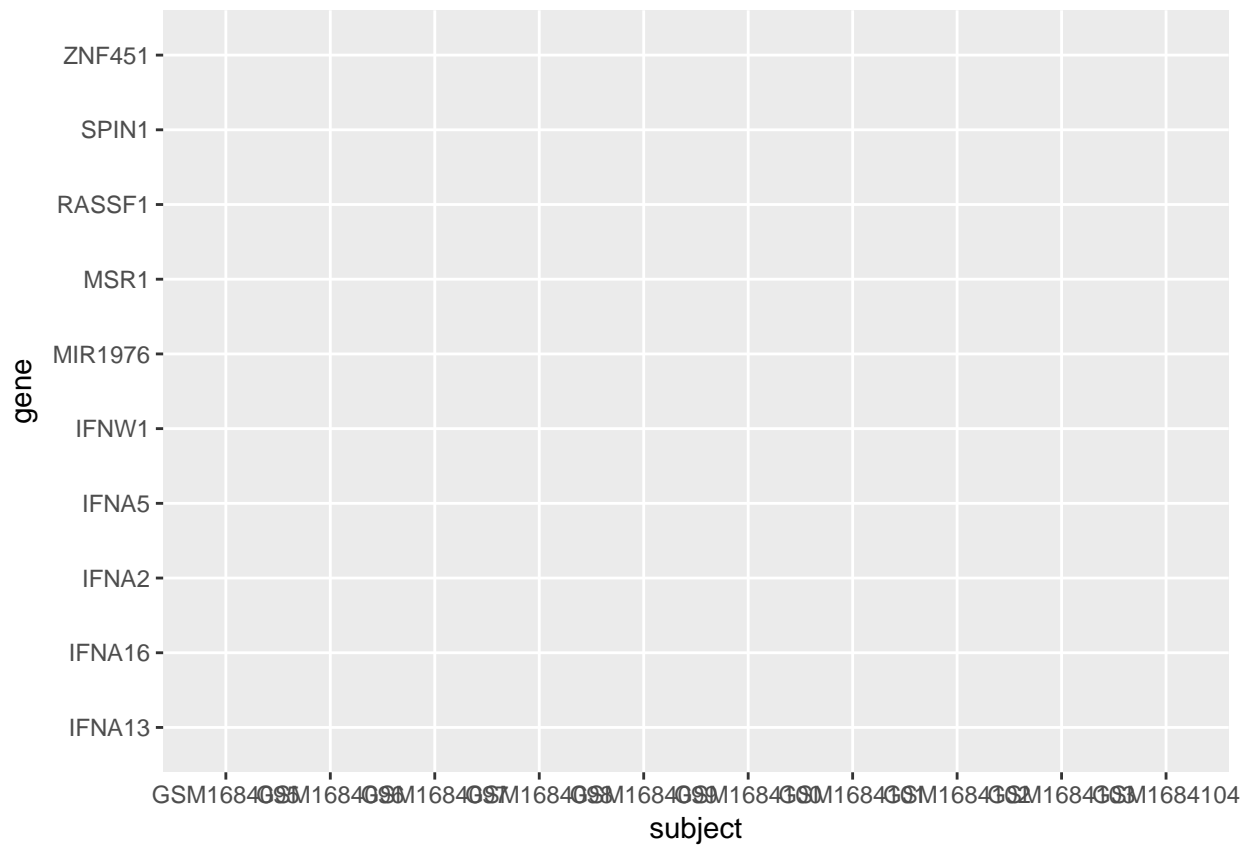
Add `geom_tile`

```r
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)

exp.long <- gather(data = exp.data,
                   key = gene,
                   value = expression,
                   -subject, -treatment)

exp.heatmap <- ggplot(data = exp.long, mapping = aes(x = subject,
                                                     y = gene,
                                                     fill = expression)) +
  geom_tile()

exp.heatmap
```
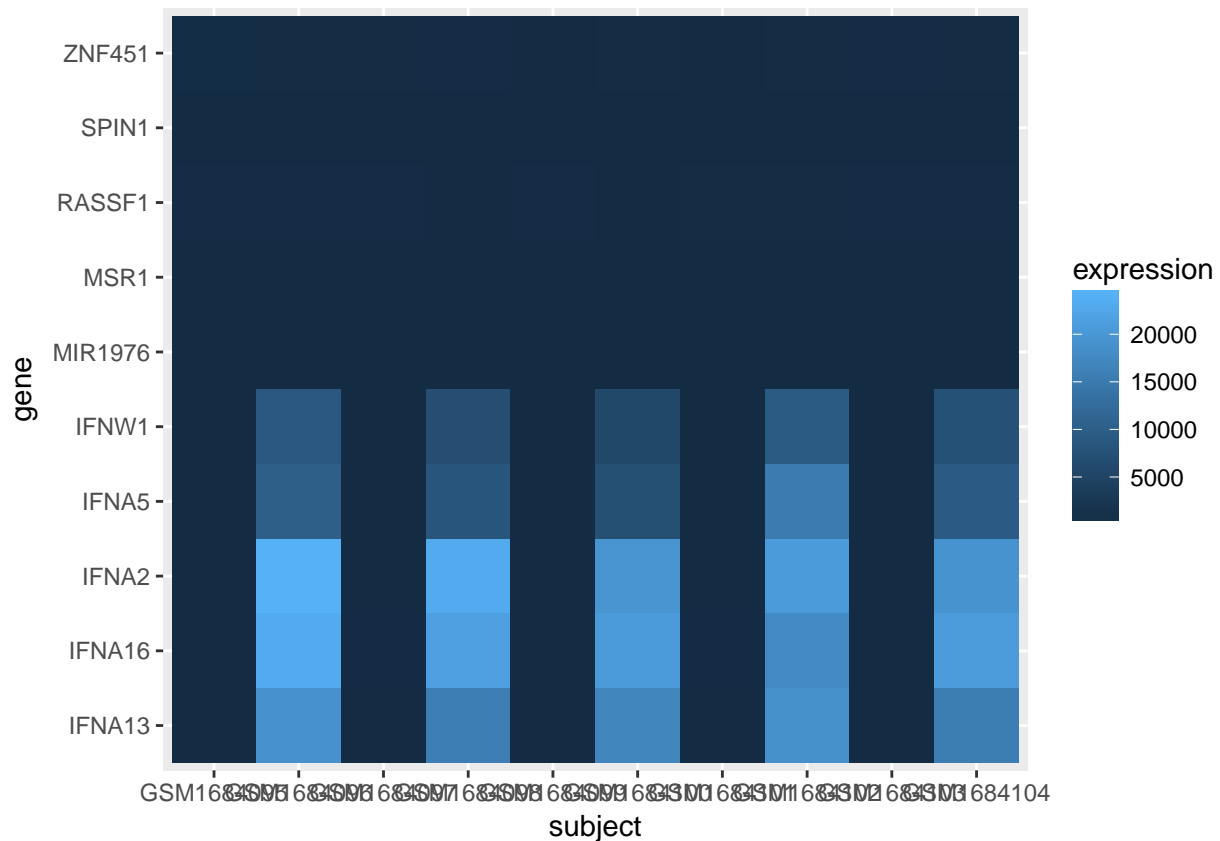
Do log transformation and graph that instead

```
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)

exp.long <- gather(data = exp.data,
                   key = gene,
                   value = expression,
                   -subject, -treatment)

exp.long$log.expression <- log(exp.long$expression)

exp.heatmap <- ggplot(data = exp.long, mapping = aes(x = subject,
                                                     y = gene,
                                                     fill = log.expression)) +
  geom_tile()

exp.heatmap
```
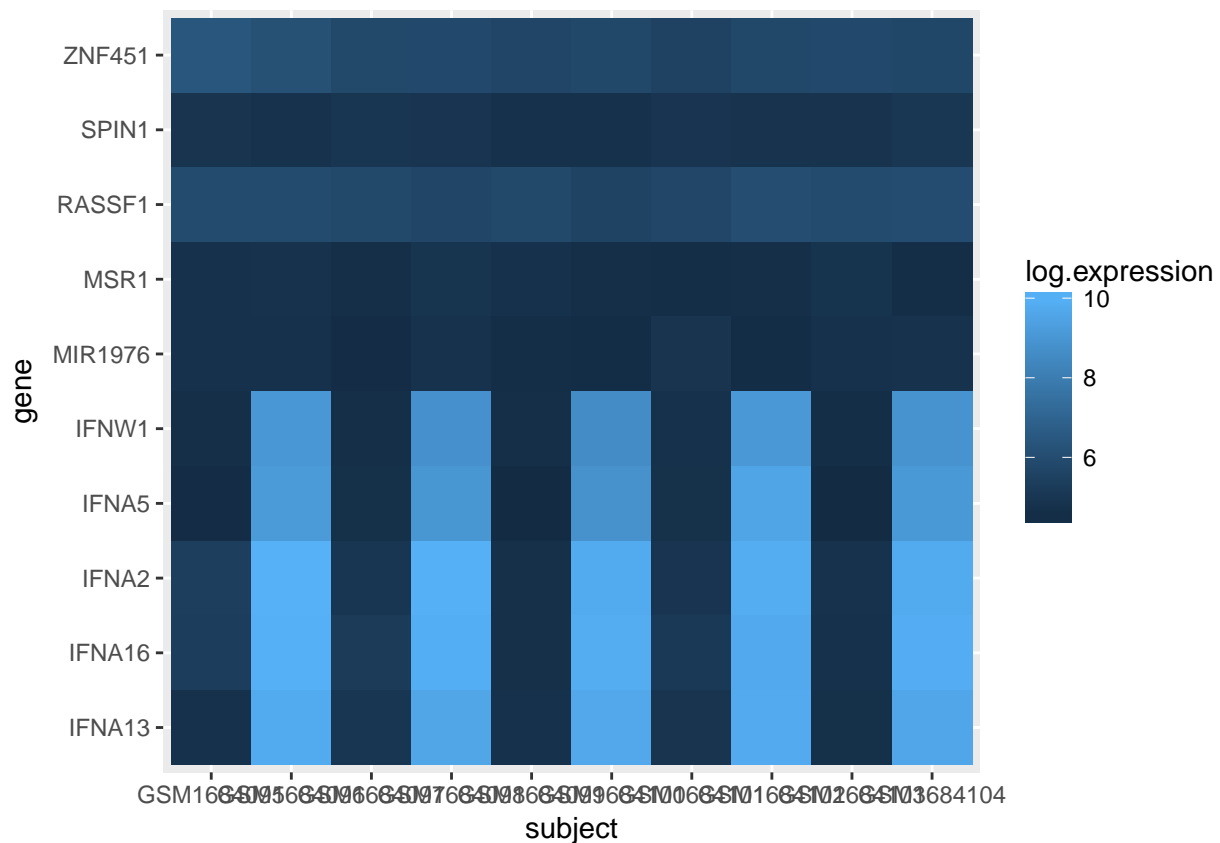
Clean up x-axis

```r
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)

exp.long <- gather(data = exp.data,
                   key = gene,
                   value = expression,
                   -subject, -treatment)

exp.long$log.expression <- log(exp.long$expression)

exp.heatmap <- ggplot(data = exp.long, mapping = aes(x = subject,
                                                     y = gene,
                                                     fill = log.expression)) +
  geom_tile() +
  xlab(label = "Subject") +
  theme(axis.title.y = element_blank(),
        axis.text.x = element_text(angle = 45, vjust = 0.5))
```
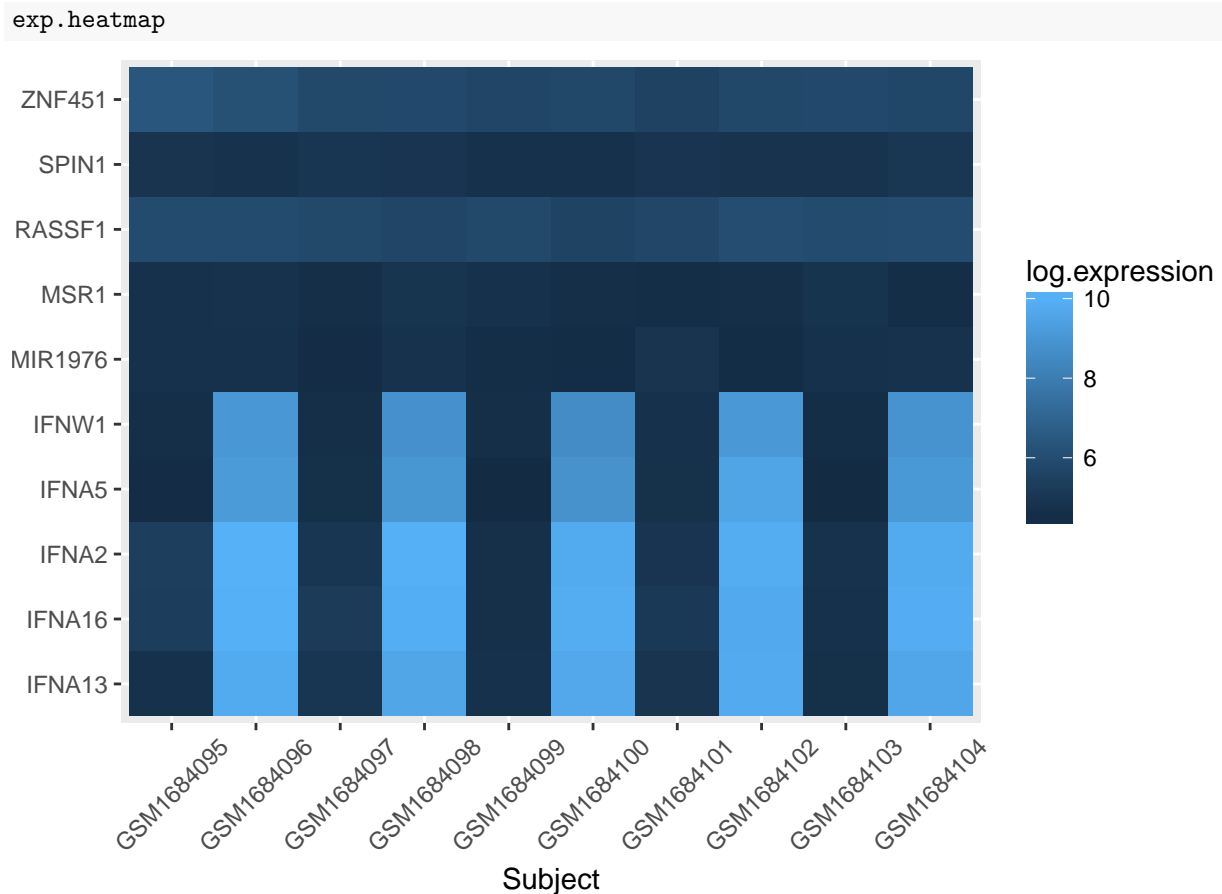
`exp.heatmap`

Add facet to separate control from flu Final (without saving)

```r
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)

exp.long <- gather(data = exp.data,
                   key = gene,
                   value = expression,
                   -subject, -treatment)

exp.long$log.expression <- log(exp.long$expression)

exp.heatmap <- ggplot(data = exp.long, mapping = aes(x = subject,
                                                     y = gene,
                                                     fill = log.expression)) +
  geom_tile() +
  xlab(label = "Subject") +
  facet_grid(~ treatment, switch = "x", scales = "free_x", space = "free_x") +
```
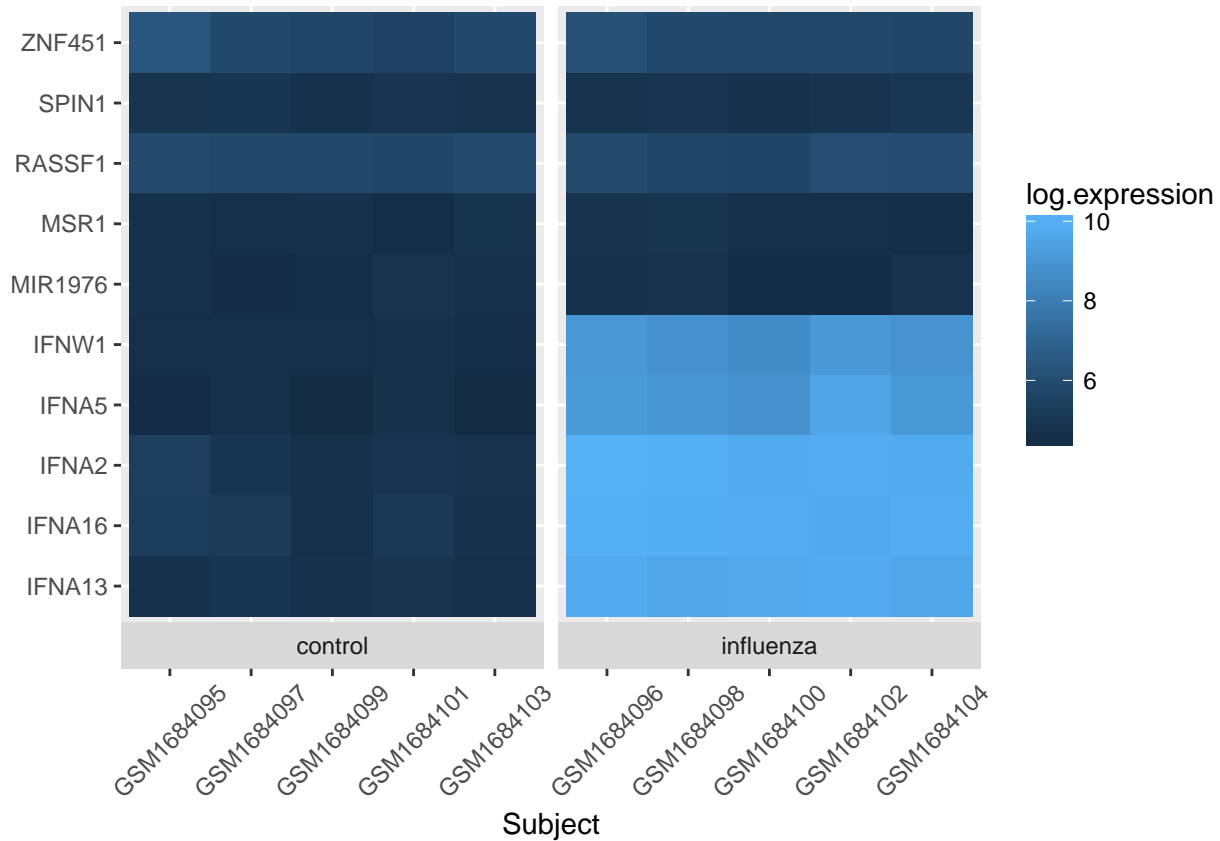
```
    theme(axis.title.y = element_blank(),
          axis.text.x = element_text(angle = 45, vjust = 0.5))
```

```
exp.heatmap
```


Final script:

```
# Expression heatmap
# Jeff Oliver
# jcoliver@email.arizona.edu
# 2017-09-14

library("tidyr")
library("ggplot2")

exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)

exp.long <- gather(data = exp.data,
                   key = gene,
                   value = expression,
                   -subject, -treatment)

exp.long$log.expression <- log(exp.long$expression)

exp.heatmap <- ggplot(data = exp.long, mapping = aes(x = subject,
                                                     y = gene,
                                                     fill = log.expression)) +
```

```
  geom_tile() +
  xlab(label = "Subject") +
  facet_grid(~ treatment, switch = "x", scales = "free_x", space = "free_x") +
  theme(axis.title.y = element_blank(),
        axis.text.x = element_text(angle = 45, vjust = 0.5))

ggsave(filename = "output/expression-heatmap.pdf", plot = exp.heatmap)
```

Additional resources

- resource one
- resource two
- A PDF version of this lesson

Back to learn-r main page

Questions? e-mail me at jcoliver@email.arizona.edu.