

Heatmaps - the gene expression edition

Jeff Oliver

15 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)

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

Create places for data and output

```
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

```
# Expression heatmap  
# Jeff Oliver  
# jcoliver@email.arizona.edu  
# 2017-09-14  
  
library("tidyr")  
library("ggplot2")
```

Read in the data

```
# 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

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

##	subject	treatment	IGFBP4	RP9
##	Length:10	Length:10	Min. : 125.7	Min. :382.1
##	Class :character	Class :character	1st Qu.: 231.1	1st Qu.:418.3
##	Mode :character	Mode :character	Median :1065.2	Median :466.9
##			Mean :1245.4	Mean :501.5
##			3rd Qu.:2031.1	3rd Qu.:588.5
##			Max. :3294.2	Max. :723.9
##	GPR34	COA4	TMEM204	UMPS
##	Min. : 142.5	Min. :143.2	Min. : 86.25	Min. :111.2
##	1st Qu.: 201.0	1st Qu.:166.4	1st Qu.: 88.33	1st Qu.:113.9
##	Median : 609.2	Median :180.5	Median : 93.35	Median :120.2
##	Mean : 709.1	Mean :185.1	Mean : 92.87	Mean :123.1
##	3rd Qu.:1034.6	3rd Qu.:208.9	3rd Qu.: 94.59	3rd Qu.:131.9
##	Max. :1664.8	Max. :225.4	Max. :105.97	Max. :143.3
##	STEAP3	IFNA10	ADAM32	POTEF
##	Min. :78.76	Min. : 104.1	Min. :80.07	Min. : 7586
##	1st Qu.:81.88	1st Qu.: 120.6	1st Qu.:82.29	1st Qu.:10479
##	Median :85.64	Median : 9009.9	Median :86.79	Median :11574
##	Mean :86.56	Mean :10605.7	Mean :87.62	Mean :12351
##	3rd Qu.:90.97	3rd Qu.:20203.3	3rd Qu.:93.64	3rd Qu.:15179
##	Max. :97.24	Max. :24029.1	Max. :95.25	Max. :17190

Transform data to long format

```
# 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 IGFBP4   367.8428
## 2 GSM1684096 influenza IGFBP4  1762.5770
## 3 GSM1684097   control IGFBP4   283.0078
## 4 GSM1684098 influenza IGFBP4  2077.6660
## 5 GSM1684099   control IGFBP4   125.7431
## 6 GSM1684100 influenza IGFBP4  1891.4320
```

Plot!

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

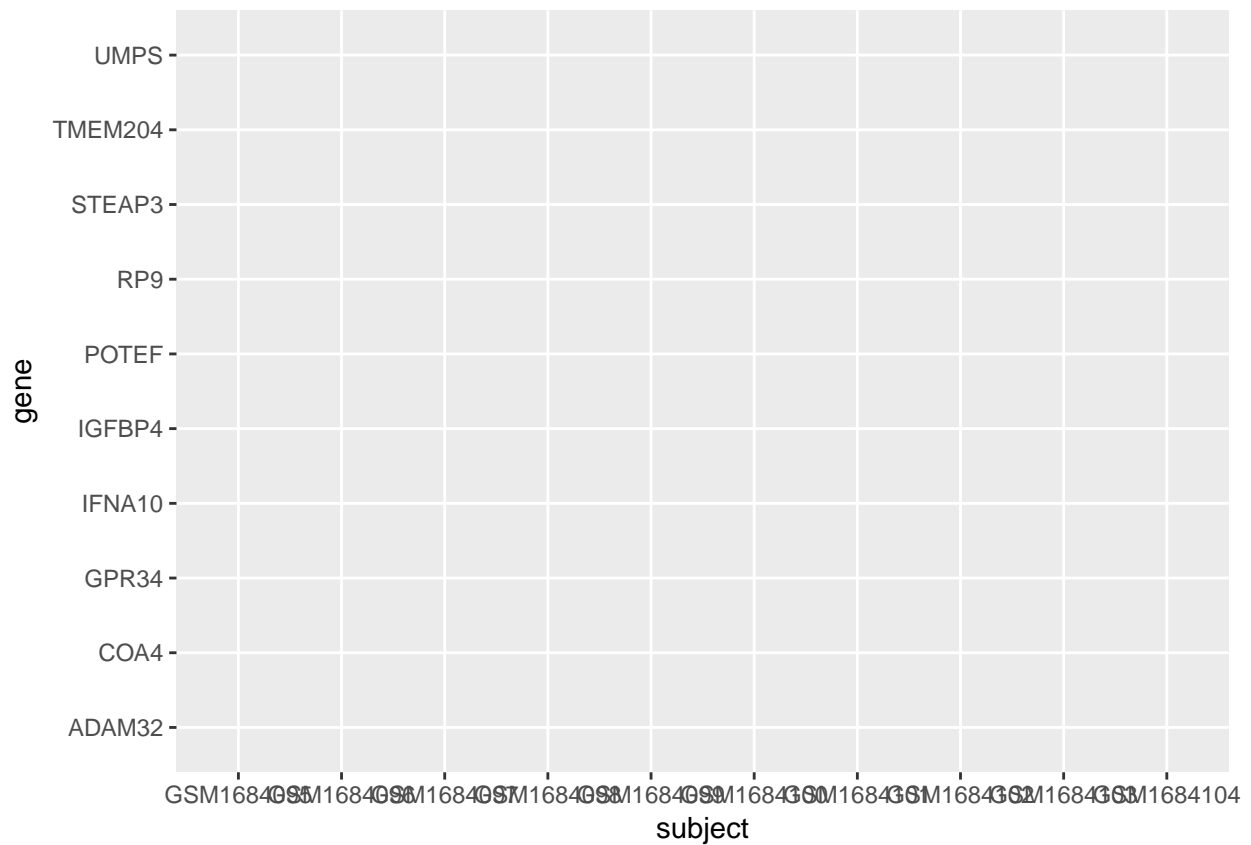
library("tidyr")
library("ggplot2")

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`

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

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

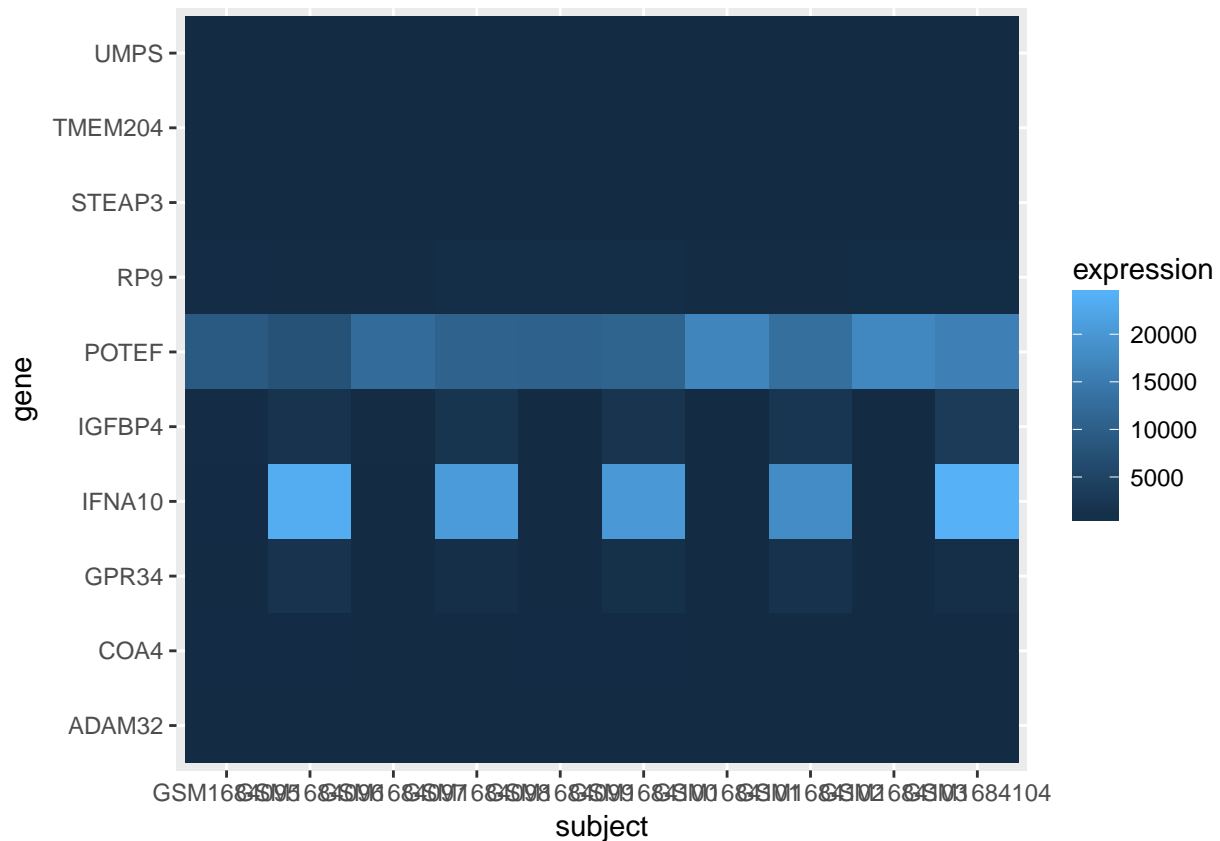
exp.data <- read.csv(file = "data/GSE68849-expression.csv", stringsAsFactors = FALSE)

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)

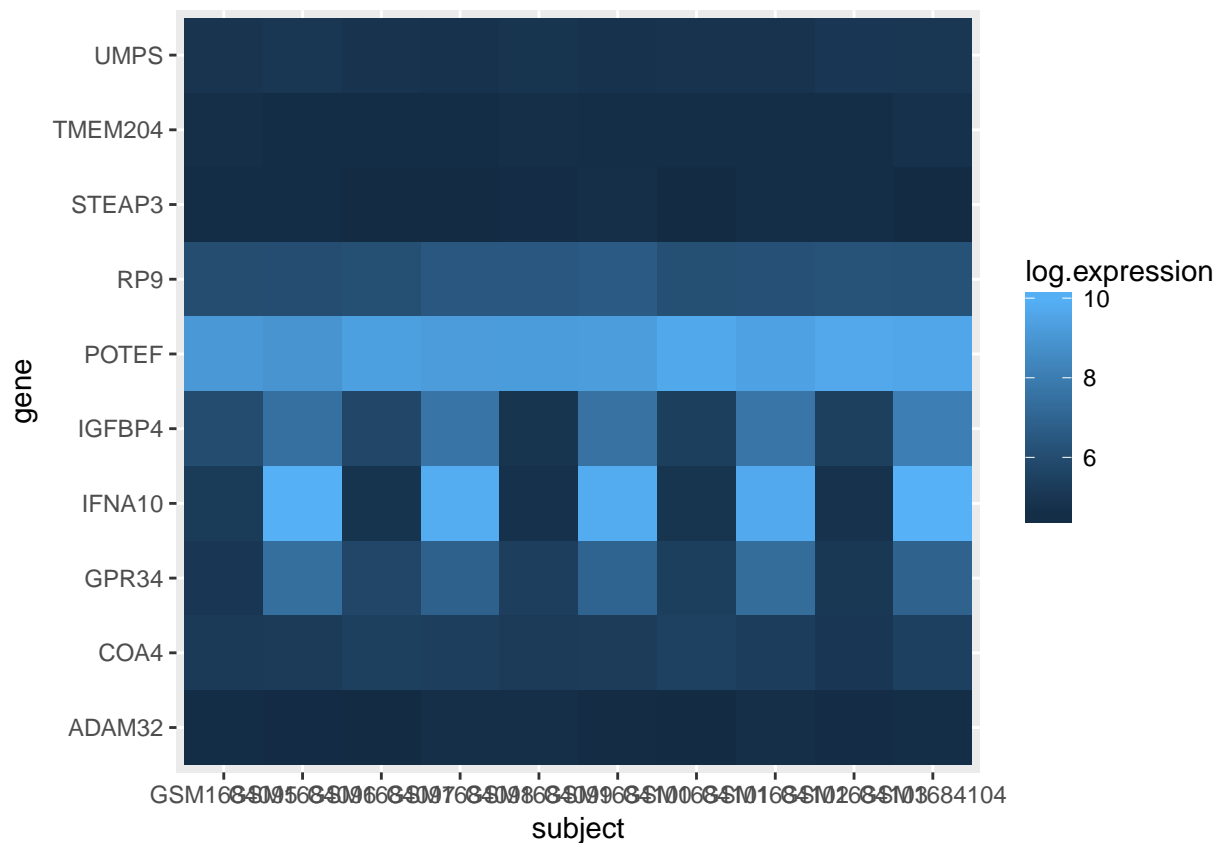
exp.long <- gather(data = exp.data,
  key = gene,
  value = expression,
  -subject, -treatment)

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

```
# 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)

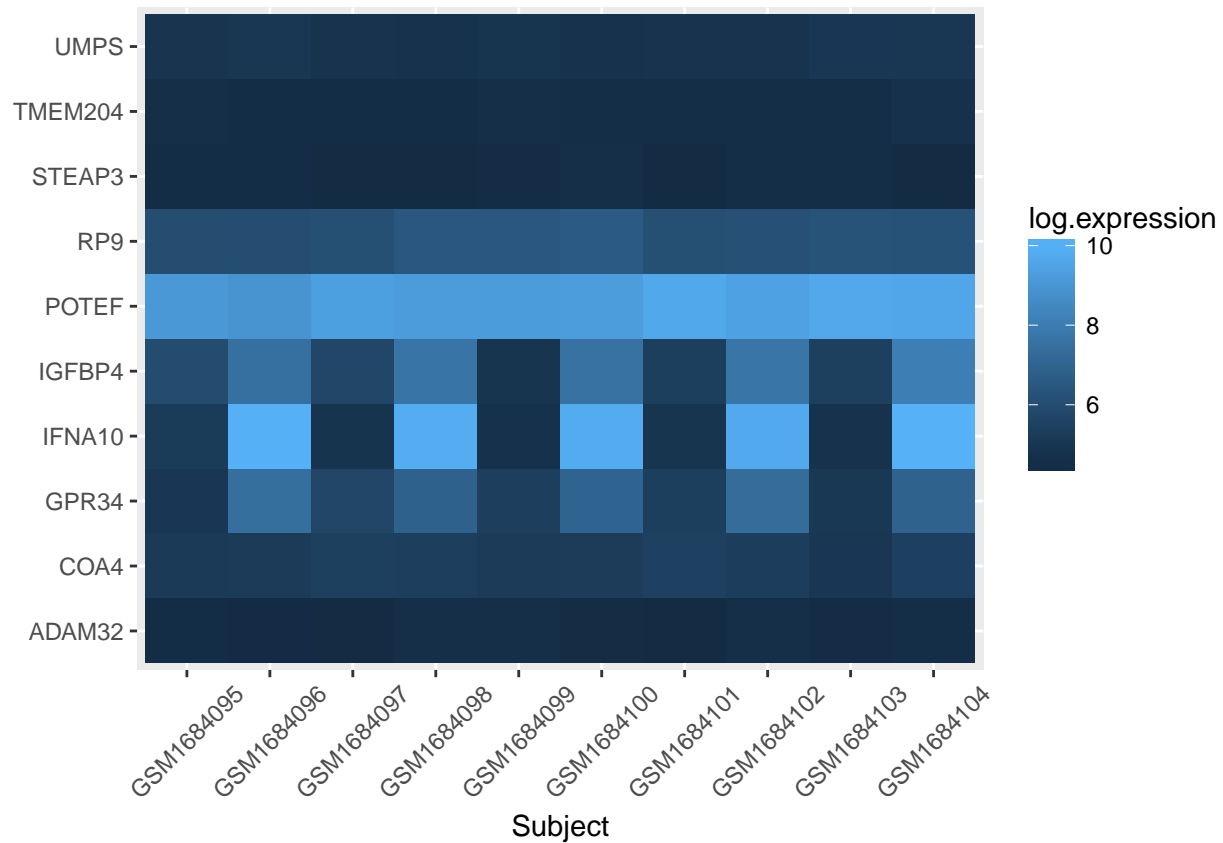
exp.long <- gather(data = exp.data,
  key = gene,
  value = expression,
  -subject, -treatment)

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)

```
# 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)

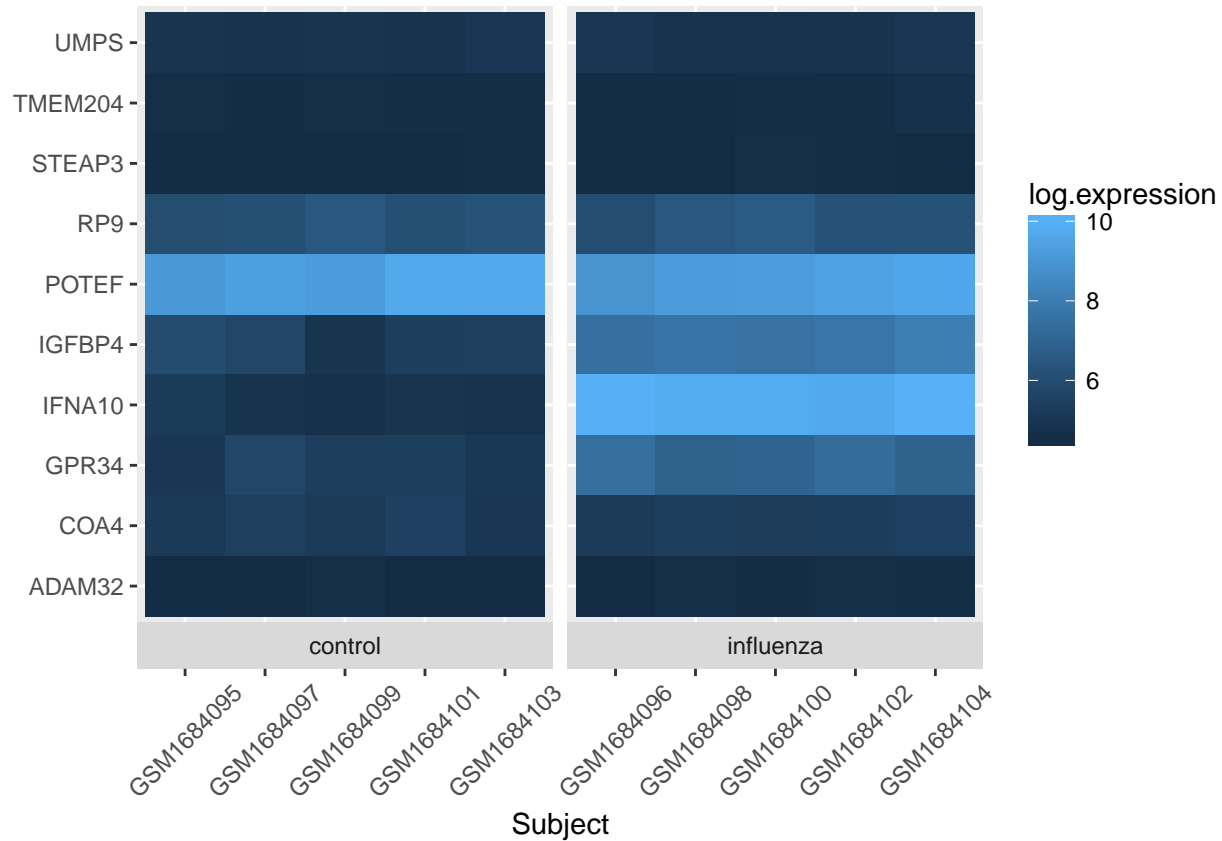
exp.long$log.expression <- log(exp.long$expression)

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.