Introduction to R Workshop

Session 1

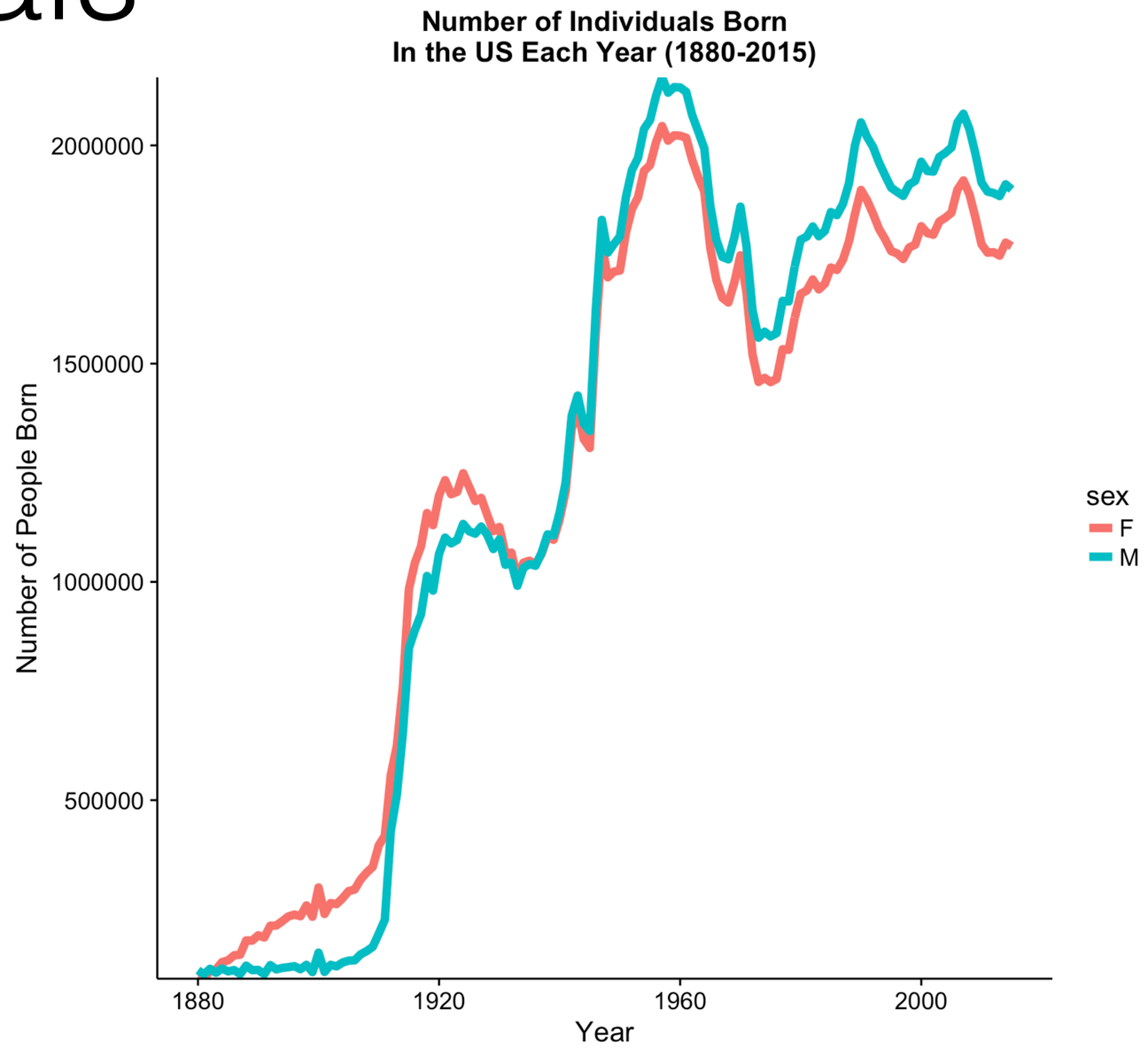
Sean Nguyen

MSU BEST

Broadening Experiences in Scientific Training

Session 1: Goals

- Install R and Rstudio
- Import packages
- Explore built in datasets
- Quick visualizations

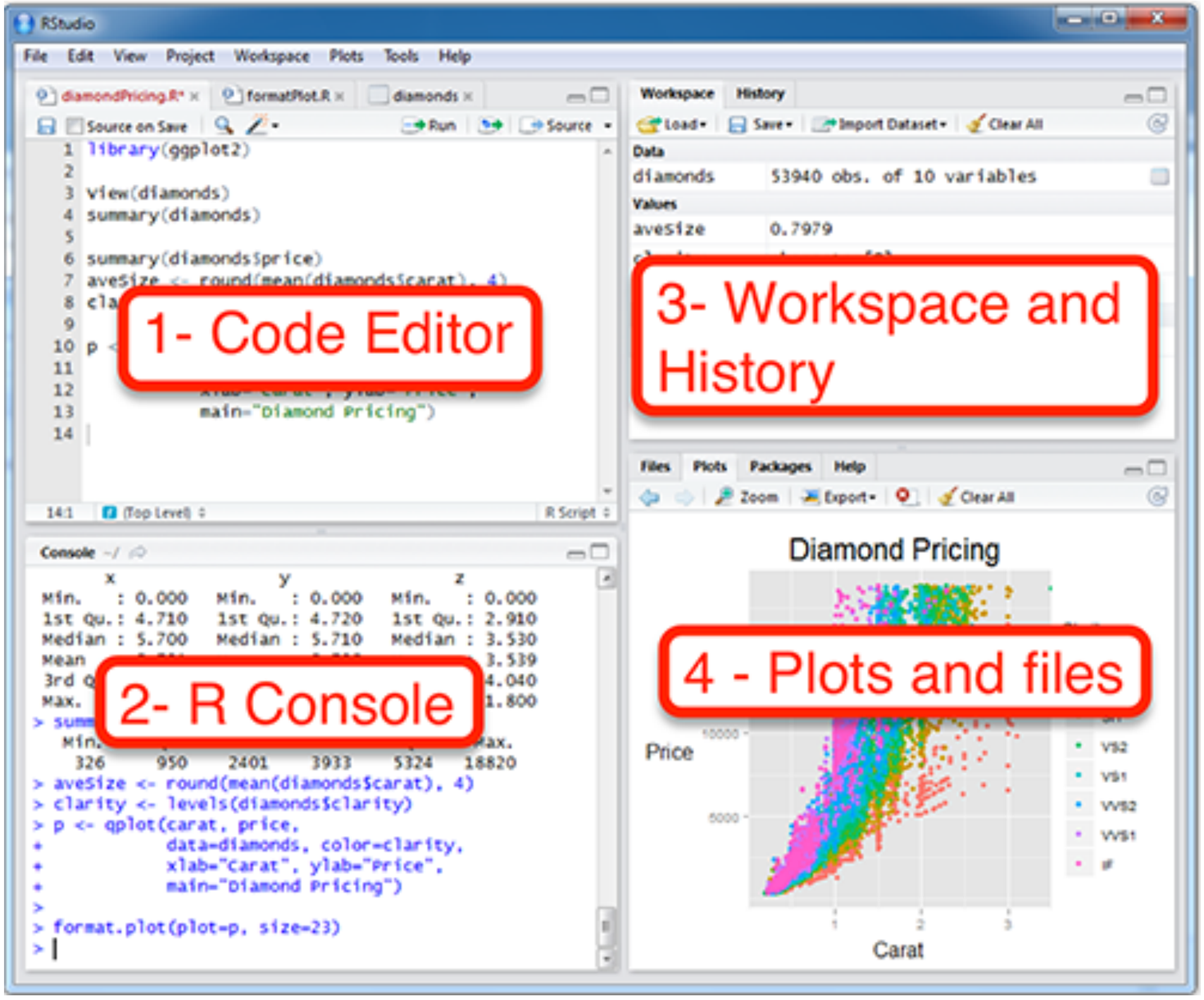


What is R?

- Programming language for statistical computing (1993)
- Reproducible analysis
  - Peer review
  - Easy to “rerun” code
- Efficient for large data analysis
- Doesn’t “manipulate” raw data

Rstudio

- Integrated development environment
- Allows for easy creation of R scripts
- More user friendly
- Great for organizing files/scripts
- Good for Rmd


RStudio
File Edit View Project Workspace Plots Tools Help
diamondPricing.R*
formatPlot.R
diamonds
Source on Save
Run
Source
1 library(ggplot2)
2
3 view(diamonds)
4 summary(diamonds)
5
6 summary(diamonds$price)
7 aveSize <- round(mean(diamonds$carat), 4)
8 cla
9
10 p <
11
12
13 main="Diamond Pricing")
14
1- Code Editor
Workspace
History
Load
Save
Import Dataset
Clear All
Data
diamonds 53940 obs. of 10 variables
Values
aveSize 0.7979
3- Workspace and History
Files
Plots
Packages
Help
Zoom
Export
Clear All
Diamond Pricing
Price
Carat
VS2
VS1
VVS2
VVS1
IF
4 - Plots and files
Console
x y z
Min. : 0.000 Min. : 0.000 Min. : 0.000
1st Qu.: 4.710 1st Qu.: 4.720 1st Qu.: 2.910
Median : 5.700 Median : 5.710 Median : 3.530
Mean 3.539
3rd Q 4.040
Max. 1.800
> summ
Min. Max.
326 950 2401 3933 5324 18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+ data=diamonds, color=clarity,
+ xlab="Carat", ylab="Price",
+ main="Diamond Pricing")
>
> format.plot(plot=p, size=23)
>
2- R Console

Install packages

Package - Collection of R functions

- Only install once
- Load them each time you run a script

tidyverse, babynames, cowplot, broom

The Assignment Operator

- Assigns value to an object

```
x <- 4
x
> 4
```

dplyr - clean up/aggregate data

- filter()
- arrange()
- select() and rename()
- mutate()
- summarize()

dplyr

Fruit	Count
Apple	34
Raspberry	67
Pear	35
Plum	27
Peach	5
Strawberry	2
Melon	97
Mango	5

filter(Fruit == “Raspberry”)

Fruit	Count
Raspberry	67

filter(Count < 10)

Fruit	Count
Peach	5
Strawberry	2
Mango	5

Pipe operator %>%

- Can be interpreted as “then”

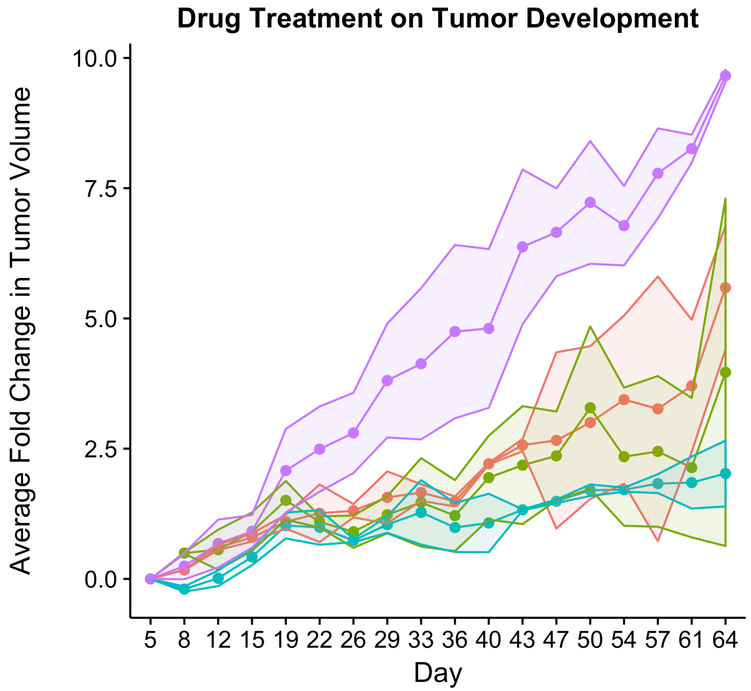
Fruit	Count
Apple	34
Raspberry	67
Pear	35
Plum	27
Peach	5
Strawberry	2
Melon	97
Mango	5

```
data %>%
    filter(Fruit == “Raspberry”)
```

Fruit	Count
Raspberry	67

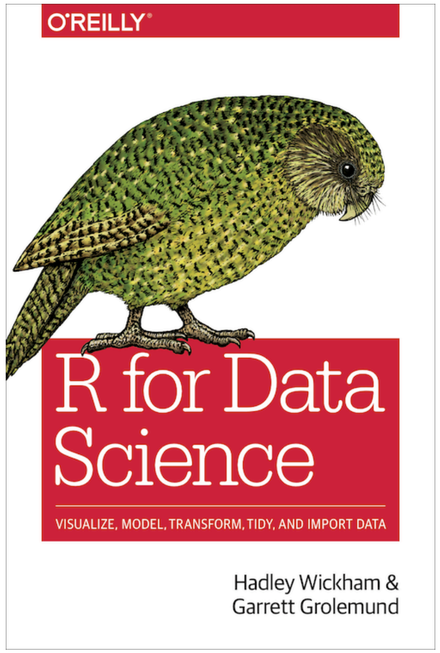
ggplot2

- Powerful graphics tool
- Generation of high quality figures
- Fully customizable



Resources

YouTube



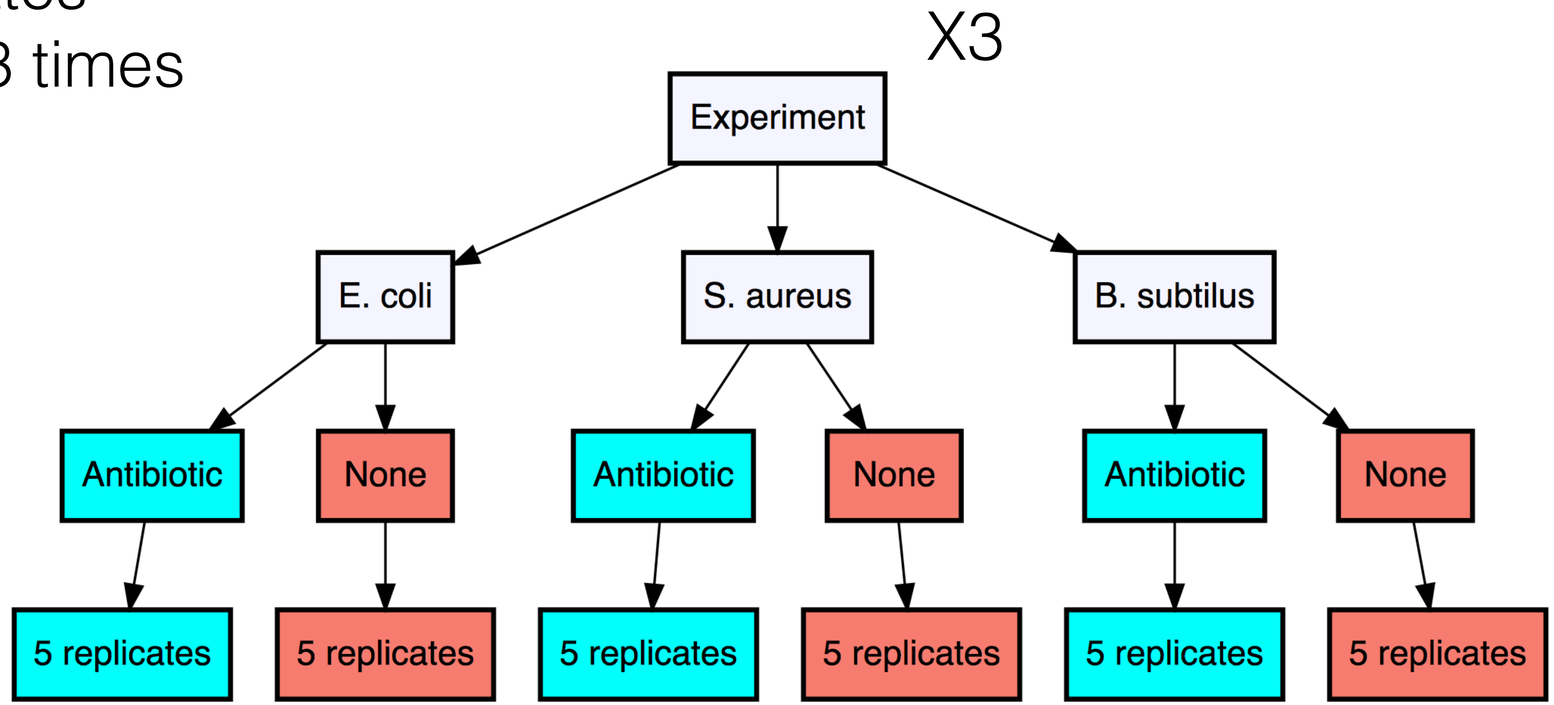
GitHub

stackoverflow

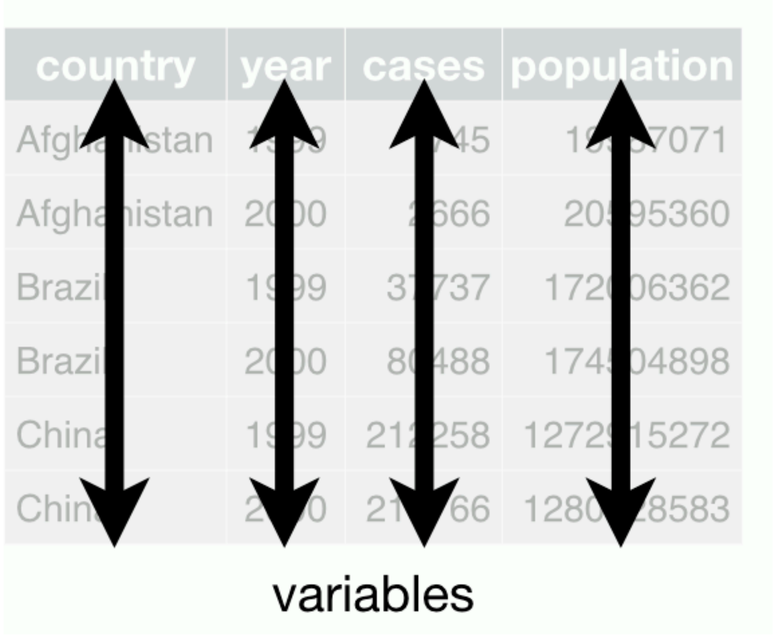
Demo!

Experimental Design

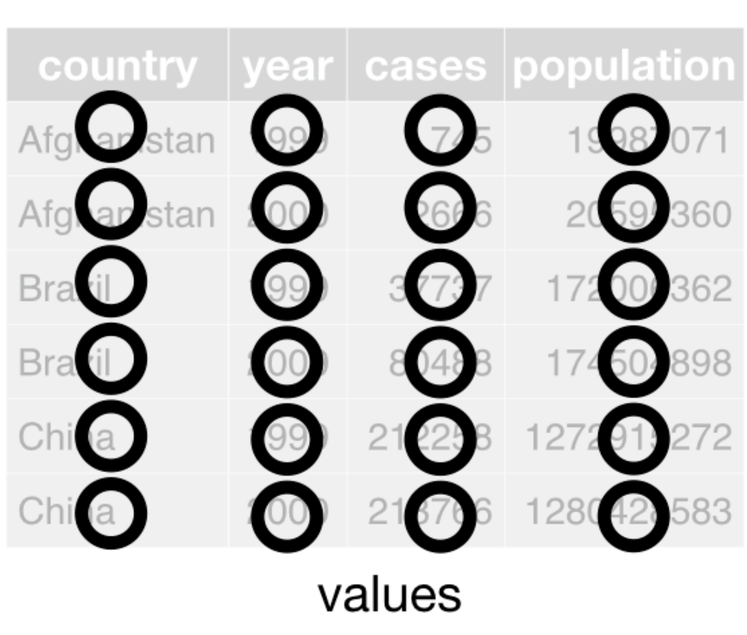
- **Three organisms** – E. coli, S. aureus, B. subtilus
- **Two treatments** - Antibiotic, None
- **Experiment** - 5 replicates
- **Repeat experiment** - 3 times



Tidy data


country
year
cases
population
variables


country
year
cases
population
observations


country
year
cases
population
values

Wide format

Treatment	1_Ecoli	1_Saureus	1_Bsubtilis	2_Ecoli	2_Saureus	2_Bsubtilis
Antibiotic	285	240	312	362	244	415
Antibiotic	345	371	461	368	375	315
Antibiotic	298	337	352	287	228	370
Antibiotic	286	394	494	378	302	314
Antibiotic	354	213	311	363	349	303
None	146	286	340	228	284	363
None	180	300	285	246	262	381
None	137	279	271	166	266	325
None	179	253	355	226	270	398
None	168	272	424	175	258	336

Long format (tidy)

Treatment	Experiment	Organism	Count
Antibiotic	1	Ecoli	285
Antibiotic	1	Ecoli	345
Antibiotic	1	Ecoli	298
Antibiotic	1	Ecoli	286
Antibiotic	1	Ecoli	354
None	1	Ecoli	146
None	1	Ecoli	180
None	1	Ecoli	137
None	1	Ecoli	179
None	1	Ecoli	168

Tidy data

Treatment	Experiment	Organism	Count
Antibiotic	1	Ecoli	285
Antibiotic	1	Ecoli	345
Antibiotic	1	Ecoli	298
Antibiotic	1	Ecoli	286
Antibiotic	1	Ecoli	354
None	1	Ecoli	146
None	1	Ecoli	180
None	1	Ecoli	137
None	1	Ecoli	179
None	1	Ecoli	168

Organism	Treatment	Experiment	N	mean	sd	se
Ecoli	Antibiotic	1	5	313.6	33.32116445	14.90167776
Ecoli	Antibiotic	2	5	351.6	36.66469692	16.39695094
Ecoli	Antibiotic	3	5	346.2	44.80736547	20.03846301
Ecoli	None	1	5	162	19.55760722	8.746427842
Ecoli	None	2	5	208.2	35.42880184	15.84424186
Ecoli	None	3	5	177.6	40.14722905	17.95438665