

# Analyze and graph your data using Prism

Zhengtao Xiao

8.26.2020

[zhengtao.xiao@duke.edu](mailto:zhengtao.xiao@duke.edu)

“The simple graph has brought more information to the data analyst’s mind than any other device.” – John Tukey

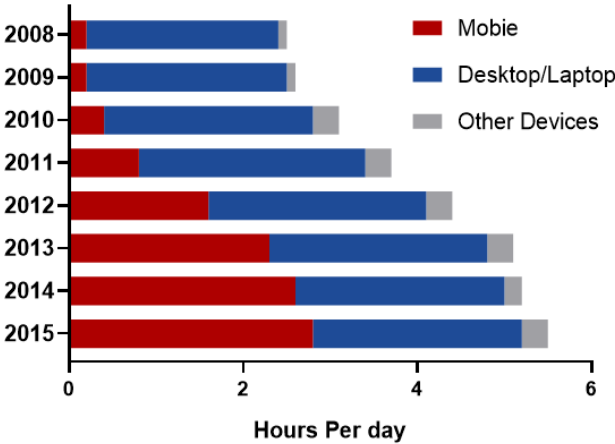
# Graph help us interpret scientific data more efficiently

Time spent with digital media

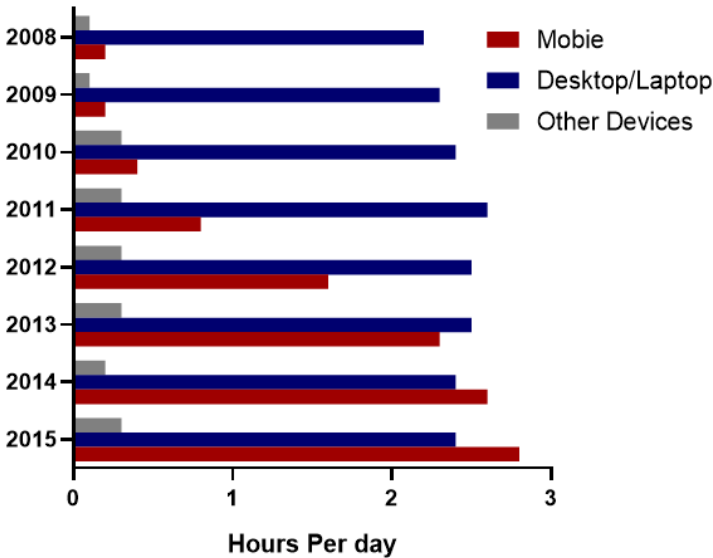
Year	Mobie	Desktop/Laptop	Other Devices
2015	2.8	2.4	0.3
2014	2.6	2.4	0.2
2013	2.3	2.5	0.3
2012	1.6	2.5	0.3
2011	0.8	2.6	0.3
2010	0.4	2.4	0.3
2009	0.2	2.3	0.1
2008	0.2	2.2	0.1

## When to use what

Time spent with digital media



Time spent with digital media



## Bad examples:

Constituent	ANOVA P-value
Oil content	0.362
$\alpha$ -Pinene	0.439
$\beta$ -Pinene	0.380
Isocaryophyllene ( $\gamma$ -Caryophyllene)	0.072
$\beta$ -Caryophyllene	0.007
$\alpha$ -(E)-Bergamotene	0.001
(Z)- $\beta$ -Farnesene	0.141
Caryophyllene oxide	0.052
Humulene epoxide 2	0.001
Selina-6-en-4-ol	0.064
Caryophylla-4(12),8(13)-dien-5 $\alpha$ -ol	0.101
Caryophylla-4(12),8(13)-dien-5 $\beta$ -ol	0.002
14-hydroxy-(Z)-Caryophyllene	0.212
$\beta$ -Bisabolol	0.001
$\alpha$ -Bisabolol	0.003
CBD	0.487
$\delta$ 9-Tetrahydrocannabinol (Dronabinol)	0.001

**Choose the most appropriate graph for your data**



<https://www.data-to-viz.com/>

## Why Graphpad Prism?

- Easy and fast
- Commonly used statistical methods
- Graphs and data are automatically updated in real time
- Reusable graph template
- ... ..

# Three steps to generate a graph

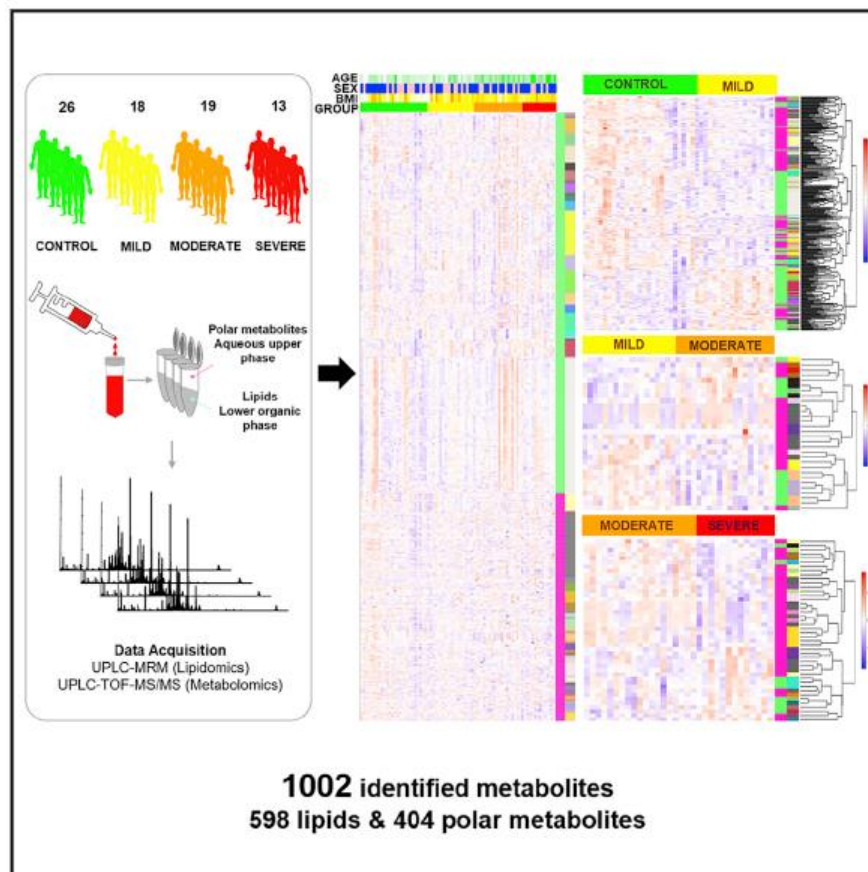
- ☐ Prepare data
- ☐ Perform analysis
- ☐ Choose the type of graph

# Dataset for practice

## Cell Metabolism Clinical and Translational Report

### Omics-Driven Systems Interrogation of Metabolic Dysregulation in COVID-19 Pathogenesis

#### Graphical Abstract



#### Authors

Jin-Wen Song, Sin Man Lam,  
Xing Fan, ..., Ji-Yuan Zhang,  
Fu-Sheng Wang, Guanghou Shui

#### Correspondence

uniquezjy@163.com (J.-Y.Z.),  
fswang302@163.com (F.-S.W.),  
ghshui@genetics.ac.cn (G.S.)

#### In Brief

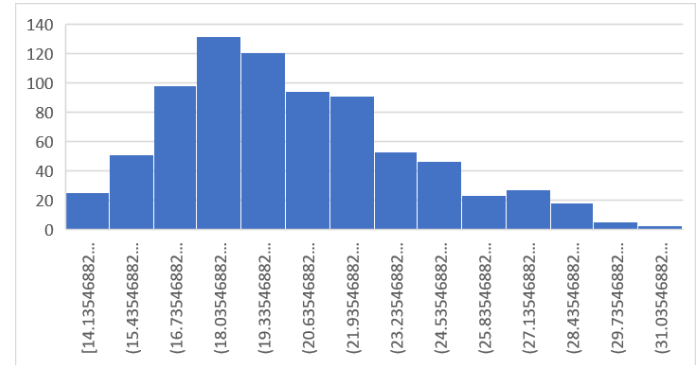
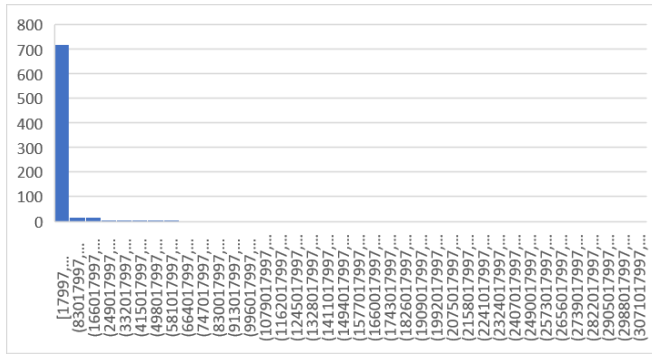
Plasma metabolite panel effectively distinguished COVID-19 patients from healthy controls (AUC = 0.975). Plasma monosialodihexosyl gangliosides (GM3s) were negatively correlated with CD4<sup>+</sup> T cell count in COVID-19 patients, and GM3-enriched exosomes were positively correlated with disease severity. These observations suggest that GM3-enriched exosomes may participate in pathological processes associated with COVID-19 progression.



- ☐ Distribution of metabolites' concentration
- ☐ Correlation of metabolomics across patients
- ☐ T-test (comparison of two groups)
- ☐ One-way anova (comparing more groups)

# Practice: Transform the data to log2 scale

1. “log2” changes the distribution to normal-like distribution



2. “log2” brings the up- and down-regulated genes/metabolites to same scale

Control = 20

Treatment = 160



Fold change = 8

Log2 fold change = 3

Control = 160

Treatment = 20

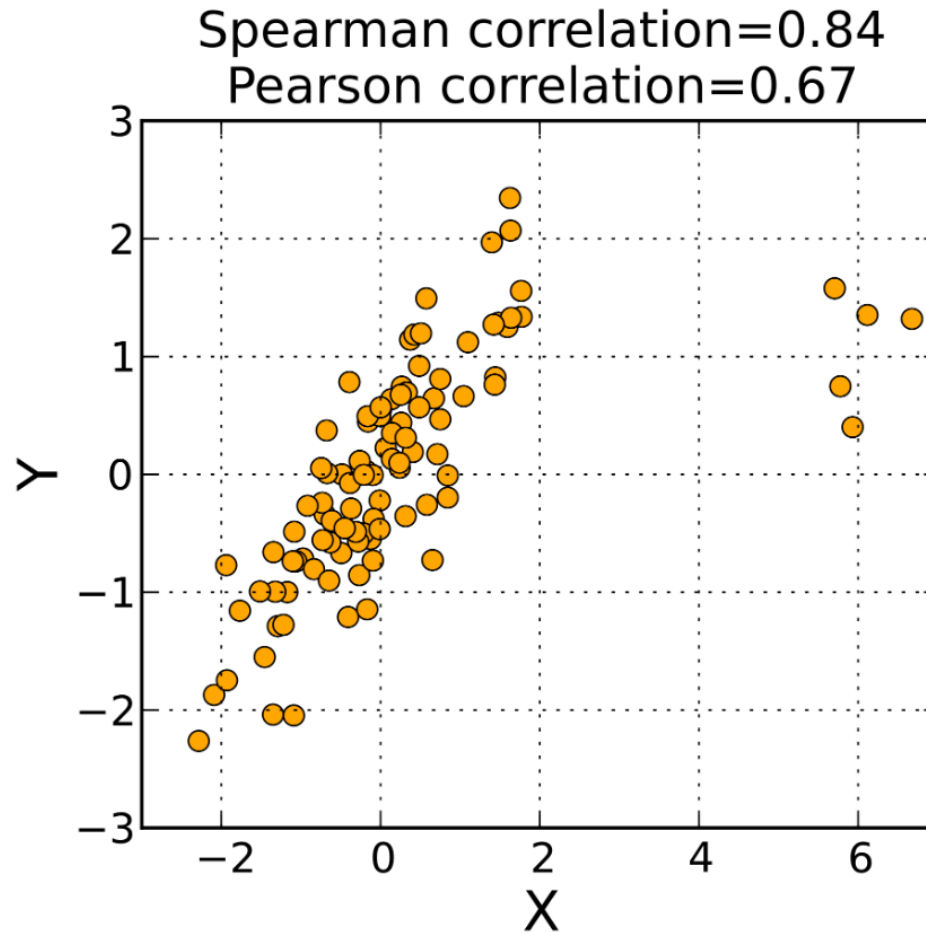


Fold change = 0.125

Log2 fold change = -3

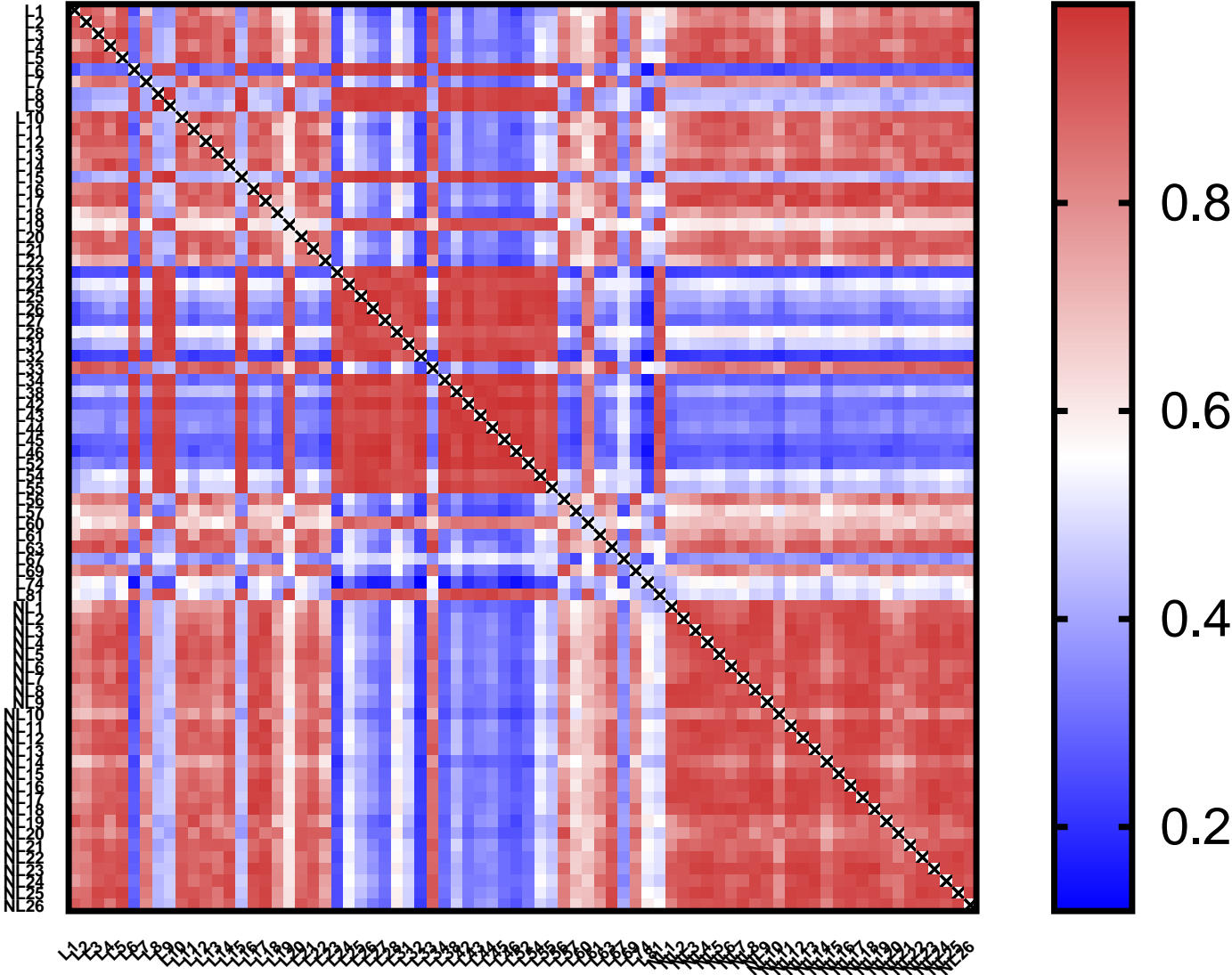
# Practice: calculate correlation

Correlation (pearson or spearman?)



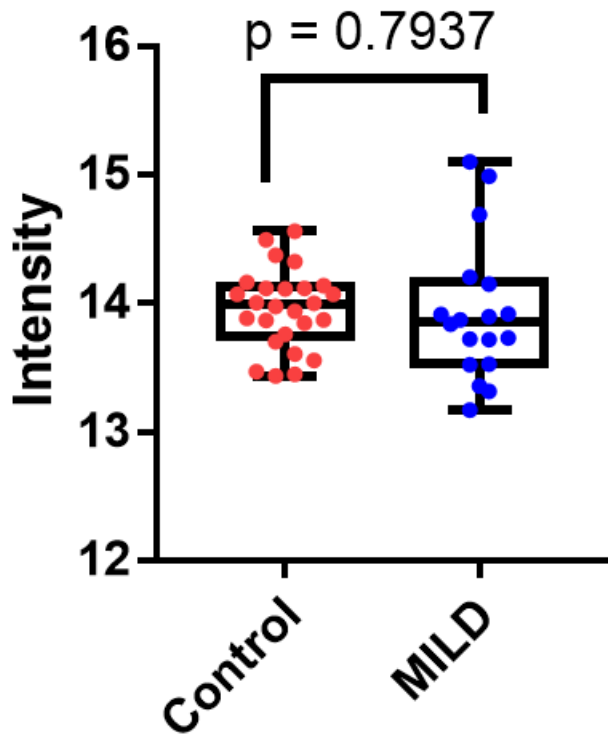
The Spearman correlation is less sensitive than the Pearson correlation to strong outliers

# Practice: correlation matrix

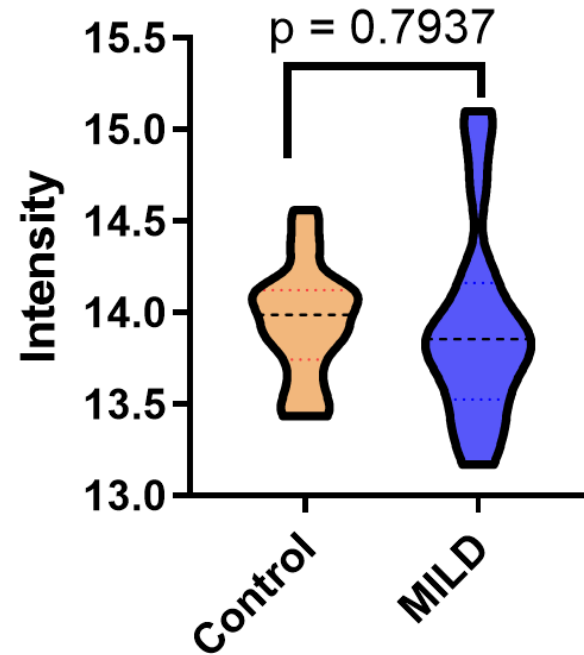


# Practice: using t-test to compare two groups

## 5-Hydroxy-L-tryptophan

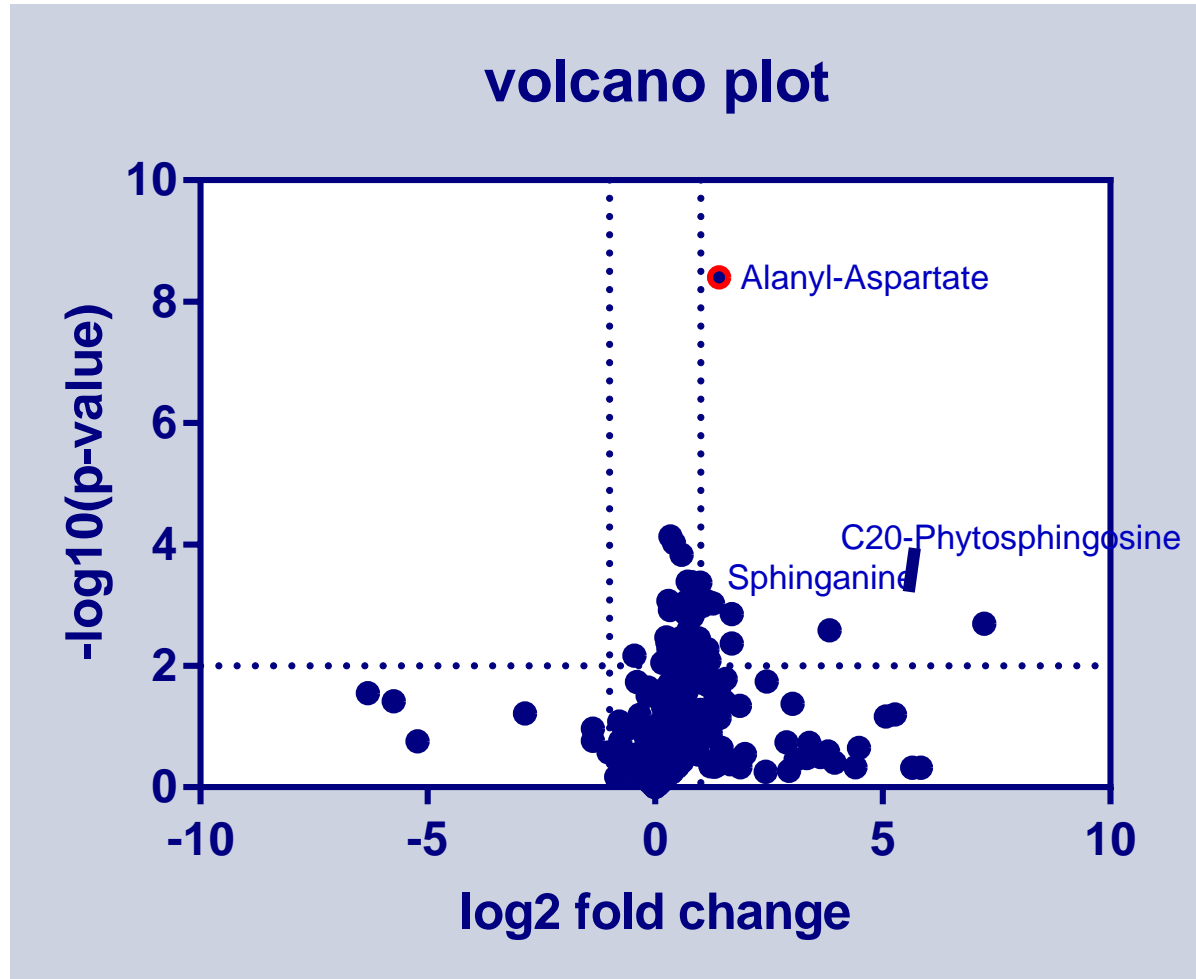


## 5-Hydroxy-L-tryptophan



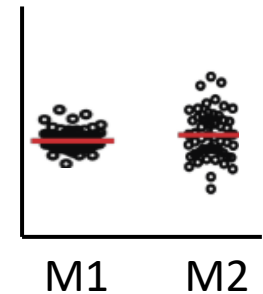
# Practice: volcano plot

Volcano plot: significance vs magnitude of changes in metabolites



## Advice: when to plot SEM vs. SD

- **SD:** standard deviation, quantifies how much the values vary from one another
- **SEM:** standard error of the mean, represents the accuracy of the true mean for the population.  $SEM = SD / (\text{square root of sample size})$
- **SD:** If you want to show the variation of your data. For example, if you want to present that one metabolite is much more stable than another one by stimulus.
- **SEM:** If you want to show how precisely you determine the true mean. For example, you want to present that one metabolite is up regulated after infection.



# Practice: using anova to compare multiple groups

We want to see if selected metabolite significantly changed after infected.

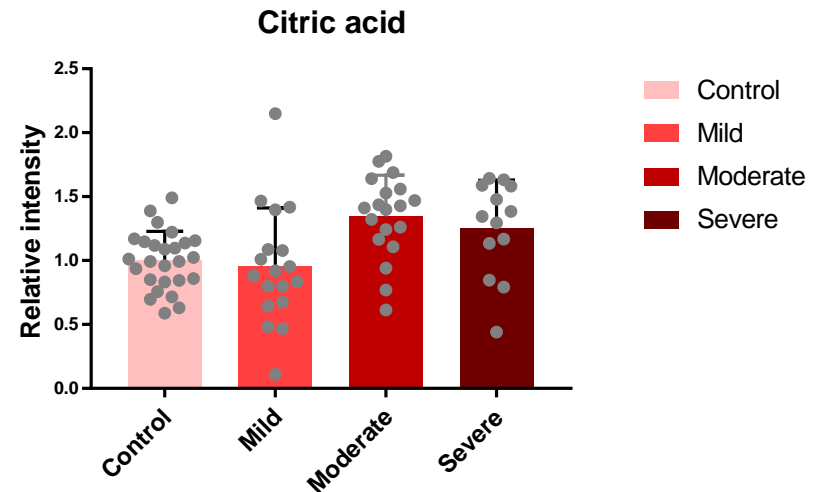
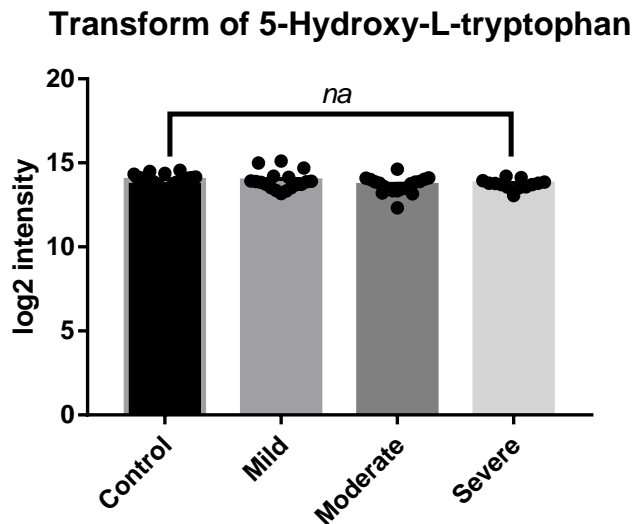
**One-way anova** (you can simply think it as a multiple t-test):

determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups

Null and Alternative hypotheses

$H_0: \mu(\text{control}) = \mu(\text{mild}) = \mu(\text{moderate}) = \mu(\text{severe})$

$H_1$ : not all  $\mu$  are equal

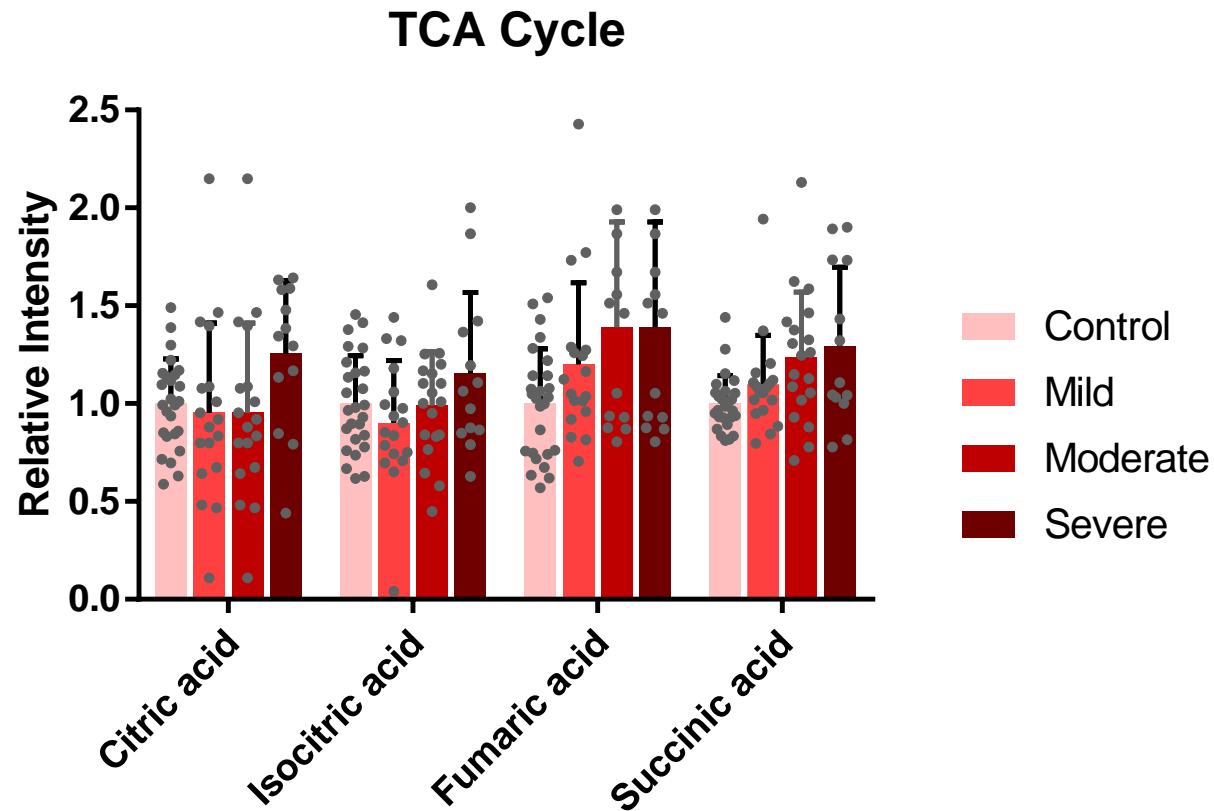


We want to see the relative intensities to control group.



# Practice: Group Bar Graph

We want to see the relative intensities of all metabolites in TCA cycle



## More information

- ☐ Prism user guide
- ☐ Prism statistics guide
- ☐ Prism Tips

zhengtao.xiao@duke.edu

## Next lecture

### An Introduction to R

---

Notes on R: A Programming Environment for Data Analysis and Graphics  
Version 3.6.1 (2019-07-05)

<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

[https://github.com/zhengtaoxiao/NCSU\\_R](https://github.com/zhengtaoxiao/NCSU_R)