

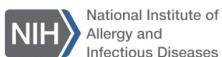
BCBB Workshop

Customizing Your Graphs

using GraphPad Prism 8

Qinlu (Claire) Wang

Bioinformatics and Computational Biosciences Branch (BCBB)
 Office of Cyber Infrastructure and Computational Biology (OCICB)
 National Institute of Allergy and Infectious Diseases (NIAID)



NIAID

1

OCICB Bioinformatics and Computational Biosciences Branch (BCBB)

- Part of NIAID
- Group of ~50
- Software developers
- Computational Biologists
- Project Management & Analysis Professionals
- Biostatistics, Genomics, Structural Biology, Proteomics, Programming
- Available to all NIAID researchers at no cost



NIAID

2

How to contact us?

1. Submit Request

Send emails to bioinformatics@niaid.nih.gov OR fill a request online at [Online Request](#)

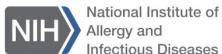
Bioinformatics and Computational Biology Request Form

COMPLETE AND SUBMIT FORM BELOW:

First Name.* <input type="text" value="Enter first name"/>	Last Name.* <input type="text" value="Enter last name"/>	Phone.* <input type="text" value="Enter phone"/>	Email.* <input type="text" value="Enter email"/>
Project Title.* <input type="text" value="Enter title"/>	Service Request.* <input type="text" value="Collaborative Data Analysis"/>		
Project Description.* <input type="text" value="Enter description"/>			

2. Attend our workshops

http://www.eventzilla.net/user/NIAID_OCICB_BCBB



NIAID

3

Outline

1. Introduction to GraphPad Prism and its features
2. GraphPad Prism data tables and data entry
3. Format single data points, specified groups, or entire data sets
4. Change the range and measurement scale of X- and Y-axes
5. Create axis breaks, add error bars, and insert special characters
6. Overlay one figure over another using Prism Layouts
7. Export graphs to MS Word, MS PowerPoint, and other options



NIAID

4

What is Prism 8?

GraphPad Prism is a versatile statistics tool for **scientists**.

1. *Structured Data Tables for scientific research*
2. *Statistical analyses and explanation*
3. *Countless ways to customize your graphs*
4. *Gain insights and guidance at every step*
5. *Collaborate with colleagues and share your research with the world.*



NIAID

5

Prism Guides

Prism provides three guides:

- [Prism User Guide](#)
- [Statistics Guide](#)
- [Curve Fitting \(Regression\) Guide](#)



NIAID

6

BCBB Prism Labs

BCBB Prism Labs:
<https://github.com/niaid/Prism>

[Lab 0: How to install Prism 8 on Mac](#)

[Lab 1: Create and edit data tables](#)

[Lab 2: Visualization](#)

[Lab 3: Descriptive Statistics](#)

[Lab 4: Statistical Testing](#)

[Lab 5: Survival Analysis](#)

[Lab 6: Categorical Data Analysis](#)

[Lab 7: Regression and Curve Fitting](#)



NIAID

7

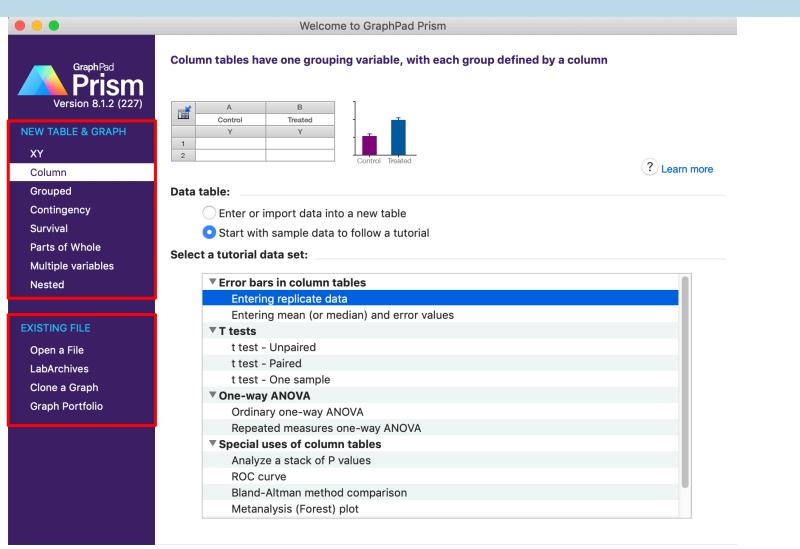
Lab 1 – Create and Edit Data Tables

1st Part: 8 Types of Data Tables

2nd Part: Read Existed Files

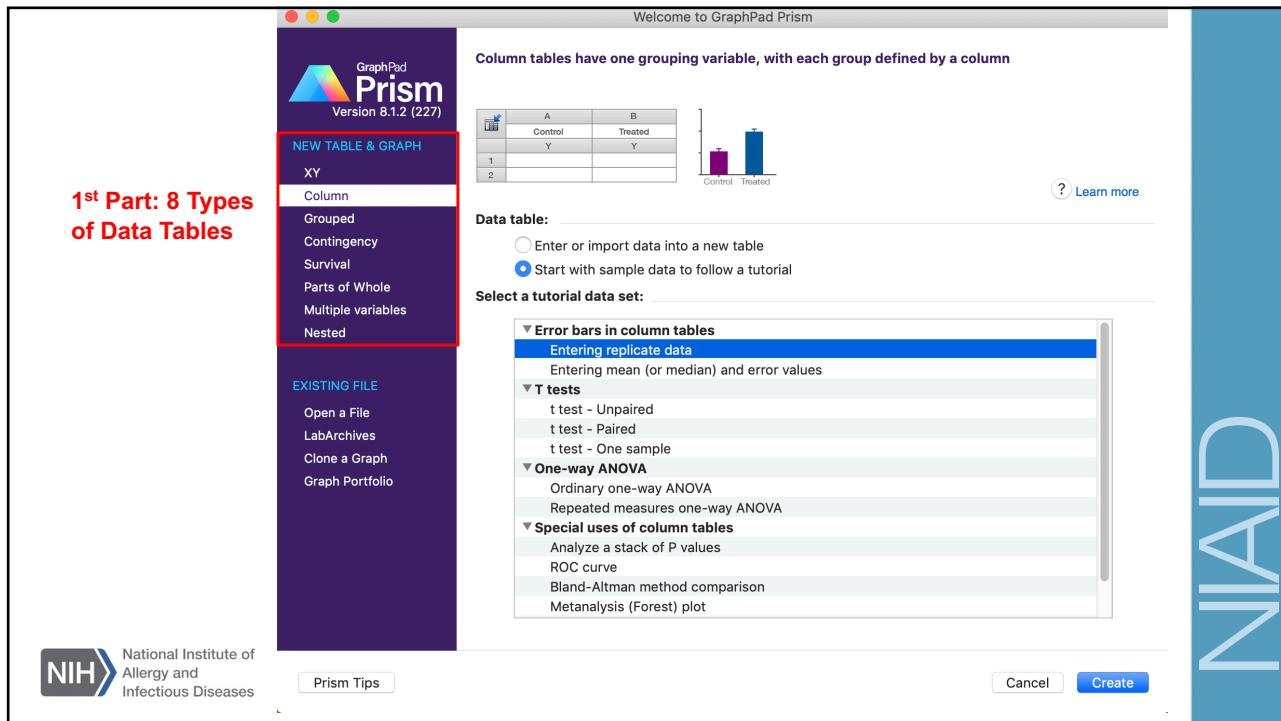


NIAID



8

4



9

Eight Types of Data Tables

- 1. *XY Tables*
- 2. *Column Tables*
- 3. *Grouped Tables*
- 4. *Contingency Tables*
- 5. *Survival Tables*
- 6. *Parts of Whole Tables*
- 7. *Multiple Variable Tables*
- 8. *Nested Tables*

NIAID

NIH National Institute of Allergy and Infectious Diseases

10

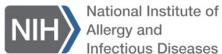
1.XY Tables

An XY table is a graph where every point is defined by both an X and a Y value. This kind of data are often fit with linear or nonlinear regression.

Analyses performed with XY data

- Nonlinear regression (curve fit)
- Linear regression
- Area Under Curve
- Deming (Model II) Linear regression
- Correlation
-

Table format: XY	X	Group A			Group B		
	Seconds	Control			Treated		
	X	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Title	1	35	31	42	36	39
2	Title	2	43	49			
3	Title	3	50	57	67	87	89
4	Title	4					62
5	Title	5	77	89	99	102	145
6	Title	10	145		121		154
7	Title	15				254	269
8	Title	20	167	187	145		231
9	Title	25				289	296
10	Title						271



11

2.Column Tables

Use column tables if your groups are defined by one scheme, perhaps control vs. treated, or placebo vs. low-dose vs. high-dose. Each column defines one group.

Analyses performed with Column data

- Unpaired / Paired t-test
- Mann-Whitney / Kolmogorov-Smirnov test / Wilcoxon test
- One-way ANOVA / Kruskal-Wallis test / Friedman test
- Descriptive Statistics
- Normality and Lognormality Tests
- Frequency Distribution
- ROC Curve
- Identify Outliers
-

	Group A		Group B	
	Placebo		Active Drug	
	Y	Y	Y	Y
1		45		67
2		23		46
3		56		113
4		76		79
5		81		123
6		87		
7		99		



12

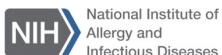
3. Grouped Tables

Grouped tables have two or more grouping variables.

Table format: Grouped		Group A			Group B		
		Control			Treated		
		A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Men	34.5	32.9	43.3	87.5	321.5*	81.7
2	Women	42.3		45.9	109.4	111.2	115.4

Analyses performed with Grouped data

- Two-way ANOVA (and mixed model)
- Three-way ANOVA (and mixed model)
- Row means with SD or SEM
- Multiple t tests - one per row



NIAID

13

4. Contingency Tables

Contingency tables are used to tabulate the actual number of subjects (or observations) that fall into the categories defined by the rows and columns of a table.

Table format: Contingency		Outcome A		Outcome B	
		Graft Patient	Graft Obstructed	Y	Y
		45	5		
1	Standard Treatment				
2	Experimental Treatment	49	1		

Analyses performed from a contingency table

- Chi-square and Fisher's exact test (also computes odds ratios and relative risk)
- Fraction of total



NIAID

14

5.Survival Tables

Survival tables are used to enter information for each subject. Prism then computes percent survival at each time, plots a Kaplan-Meier survival plot, and compares survival with some tests.

Analyses performed with survival data

- Kaplan-Meier
- Log-rank test
- Wilcoxon-Gehan test

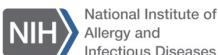


	Table format: Survival	X	Group A	Group B	Group C
		Days	Control	Treatment A	Treatment B
	X	Y	Y	Y	
1	CO	78	1		
2	NT	34	1		
3	RO	123	0		
4	LT	45	1		
5	RE	234	1		
6	AT	345	1		
7	ME	123	1		
8	NT	211	1		
9	WO	356	1		
10	RK	378	0		
11	PL	88		1	
12	AC	321		1	
13	EB	211		1	
14	OT	111		0	
15	RE	156		0	
16	AT	178		1	
17	ME	236		1	
18	NT	198		1	
19	XX	211		1	
20	XY	234		1	
21	BO	257			1
22	OW	322			1
23	HO	344			1
24	TO	365			1
25	YO	245			0
26	UT	299			1
27	OO	351			1

15

NIH NIAID

6.Parts of Whole Tables

A Parts of whole table is used when it makes sense to ask: What **fraction of the total** is each value? This table is often used to make a pie chart.

Analyses performed on parts of whole data

- Fraction of total
- Chi-square goodness of fit (compare observed distribution with theoretical distribution)



	Table format: Parts of whole	A	Number of Students
		Y	
1	A		23
2	B		29
3	C		7
4	D		2
5	E		0

16

NIH NIAID

7. Multiple Variable Tables

A multiple variable table is arranged the same way most statistics programs organize data.

Table format: Multiple variables	Variable A	Variable B	Variable C	Variable D	Variable E	Variable F	Variable G	Variable H	Variable I	Variable J
	Glycosylated hemoglobin %	Total cholesterol	Glucose	HDL	Age in years	Male?	Height in inches	Weight in pounds	Waist in inches	Hip in inches
1 Title	4.309999943	203	82	56	46	0	62	121	29	38
2 Title	4.440000057	165	97	24	29	0	64	218	46	48
3 Title	4.639999866	228	92	37	58	0	61	256	49	57
4 Title	4.630000114	78	93	12	67	1	67	119	33	38
5 Title	7.719999790	249	90	28	64	1	68	183	44	41

Analyses performed on multiple variable data

- Correlation matrix
- Multiple linear regression
- Transform and select
- Identify outliers
- Descriptive statistics
-



NIAID

17

8. Nested Tables

A nested table is used when there are two levels of nested or hierarchical replication.

Group A			Group B			Group C		
Traps + Pour on			Pour on			No vector control		
Herd 1	Herd 2	Herd 3	Herd 4	Herd 5	Herd 6	Herd 7	Herd 8	Herd 9
28	32	27	25	26	25	21	19	18
26	27	25	24	28	26	19	18	20
27	28	29	27	29	24	17	23	19
31	29	27	23	27	23	20	20	18

Analyses performed from nested tables

- Nested *t* test
- Nested one-way ANOVA
- Descriptive statistics
- Normality and lognormality tests
- Outlier tests
- One-sample *t* test



NIAID

18

Quiz: What kind of data table is it?

Table format:		Group A	Group B	Group C	Group D
		Control	Treatment 1	Treatment 2	Treatment 3
1	GS	x	Y	Y	Y
2	JM		23	34	37
3	HM		45	65	99
4	DR		31	33	36
5	PS		15	25	30
					26

1st Data Table

- The four columns represent four sequential treatments.
- Each row represents a different subject



19

NIAID

Quiz: What kind of data table is it?

Table format:		Group A			Group B			Group C			Group D		
		Low fat Male			Low fat Female			High fat Male			High fat Female		
	x	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3	C:Y1	C:Y2	C:Y3	D:Y1	D:Y2	D:Y3
1	Light smoker	24.1	29.2	24.6	20.0	21.9	17.6	14.6	15.3	12.3	16.1	9.3	10.8
2	Heavy smoker	17.6	18.8	23.2	14.8	10.3	11.3	14.9	20.4	12.8	10.1	14.4	6.1

2nd Data Table

This example examines factors that affect exercise tolerance: gender, low vs. high fat diet, and light vs. heavy smoker.



NIAID

20

Quiz: What kind of data table is it?

3rd Data Table

- The X column records time.
- The response at each time point is entered in triplicate for Control and Treated conditions.
- Some cells are blank to denote missing data.

	Table format: X	X	Group A			Group B		
		Minutes	Control			Treated		
	X	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3	
1	Title	1.0	8887	7366	9612	6532	7905	7907
2	Title	2.0	8329		8850	5352	5841	6277
3	Title	3.0	7907	8810	8669	5177	4082	3157
4	Title	4.0	7413	8481	6489	3608		4226
5	Title	5.0	7081	7178	5716	2559	3697	2816
6	Title	6.0	6249	6492		1671	3053	2891
7	Title	8.0	5442	6172	6409	2264	1658	1879
8	Title	10.0	4020	3758	4138	1905	1302	1406
9	Title	14.0	4559	3146	2547	2994	1338	739
10	Title	20.0	3033	6587	2754	1444		760
11	Title	25.0	2105	1707	2152	281	484	765
12	Title	30.0	1005	2156	1185	1103	1517	833
13	Title	50.0	820	1513	1591	1918	1128	1293



21

Quiz: What kind of data table is it?

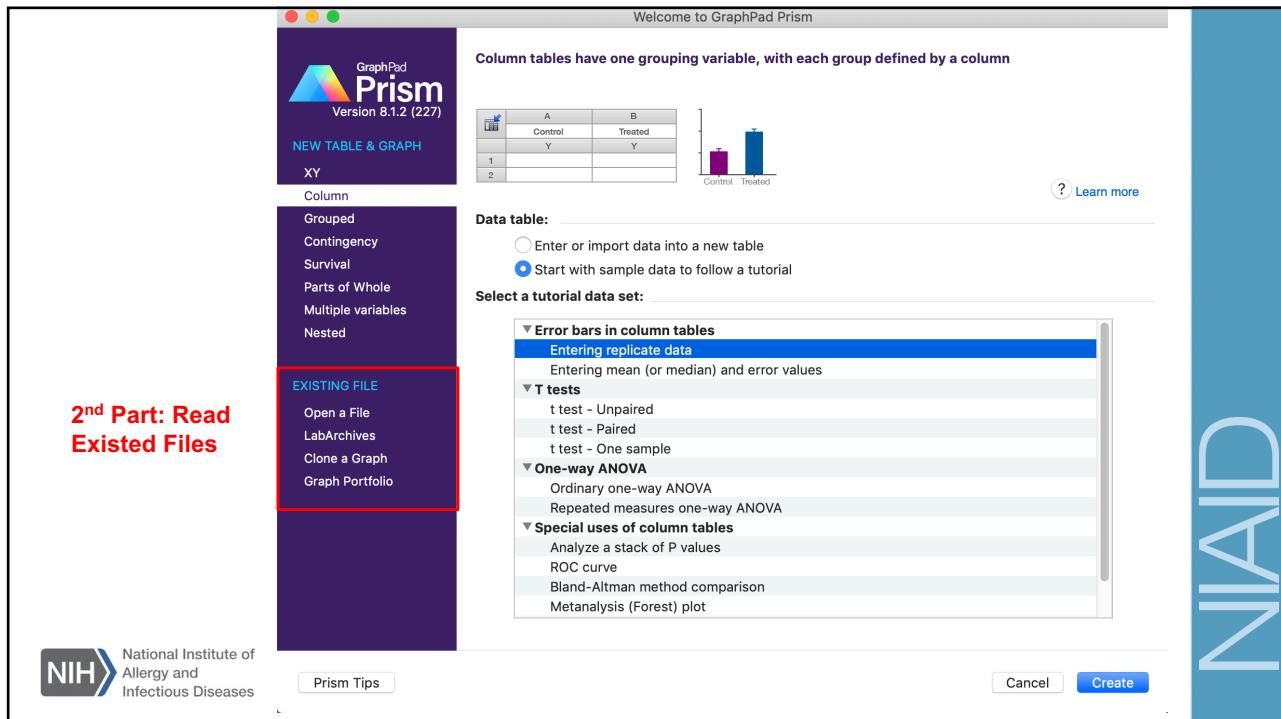
4th Data Table

- Each row represents one subject.
- The X values are time.
- The Y values are entered into three columns that define the treatment groups
- The Y value is "1" when the subject died at the specified time, and "0" when the subject's data was censored at that time

	Table format: X	X	Group A	Group B	Group C
		Days	Control	Treatment A	Treatment B
		X	Y	Y	Y
1	CO	78	1		
2	NT	34	1		
3	RO	123	0		
4	LT	45	1		
5	RE	234	1		
6	AT	345	1		
7	ME	123	1		
8	NT	211	1		
9	WO	356	1		
10	RK	378	0		
11	PL	88		1	
12	AC	321		1	
13	EB	211		1	
14	OT	111		0	
15	RE	156		0	
16	AT	178		1	
17	ME	236		1	
18	NT	198		1	
19	XX	211		1	
20	XY	234		1	
21	BO	257			1
22	OW	322			1
23	HO	344			1
24	TO	365			1
25	YO	245			0
26	UT	299			1
27	OO	351			1



22



23

The 5 Sections of a Prism Project

Data Tables

- Unpaired t test data
- New Data Table...

Info

- Project info 1
- New Info...

Results

- Unpaired t test of Unpaired t test data
- New Analysis...

Graphs

- Unpaired t test data
- New Graph...

Layouts

- Layout 1
- New Layout...

How the data are organized

The two columns define two groups. Note that, unlike many statistics programs, Prism does not define groups using a grouping variable. Instead, the groups are defined by columns.

The goals

- To determine if the differences between the two group means is greater than you'd expect to see by chance.
- To determine the 95% confidence interval for the difference between the two means.

How to perform an unpaired t test

Click Analyze, choose t test from the list of column analyses, then choose an unpaired t test on the dialog. Click the link below for detailed instructions.

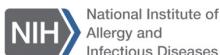
Step-by-step instructions for performing an unpaired t test

24

Two Formats

When you save a Prism file, you save the entire project into one file. You have a choice of two file formats:

- **PZF format:** This is a binary format that can be opened by Prism 4 or later, but not by other applications.
- **PZFX format:** This is a format that can be opened only by Prism 5 or later. The first part of the file contains all the data tables and info sheets in a plain-text XML format that can be viewed by other programs. After that comes information about results, graphs and layouts in a format that is incomprehensible to any program but Prism.



NIAID

Lab 2 – Visualization

- How to create graphs in Prism
- Eight ways to change a graph
- Rotating, flipping, reversing
- Changing a graph's shape and size
- Breaking Y axis into two segments
- Overlaying two column graphs
- Graphing Example: XY plots – Nonlinear curve fitting
- Exporting images from Prism



NIAID

Exporting images from Prism

Export Formats	PDF	TIFF / TIF
Advantages	File can be stretched to any size with no loss of quality	Tend to be trouble-free when submitted to journals
Disadvantages	1. Few Windows programs import pdf images 2. Few Journals accept them.	You need to choose resolution and size



NIAID

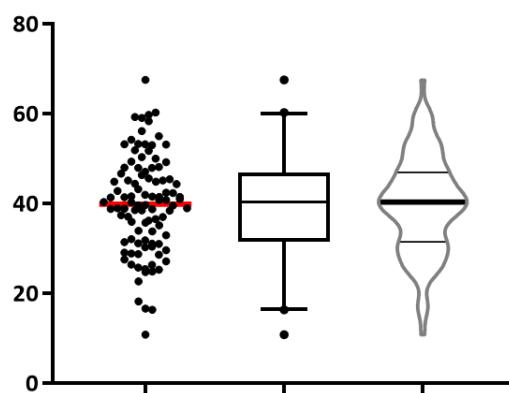
27

Viewing Data Distributions

Left: Scatter Plot

Middle: Box-and-Whiskers Graph

Right: Violin Plot



NIAID

28

Lab 5 – Survival Analysis

In many clinical and animal studies, the outcome is survival time. The goal of the study is to determine whether a treatment changes survival. Prism creates survival curves, using the product limit method of Kaplan and Meier, and compares survival curves using both the **log-rank test** and the **Gehan-Wilcoxon test**.

Examples of **events** and **censor**:

Events	Censor
1. Death	1. Still alive at the end of the study
2. Targeted Events	2. Drop out of the study



29

NIAID

Log-rank Test

- Gives equal weights to all time points
- More standard
- More powerful if the assumptions of proportional hazards** is true.

Gehan-Breslow-Wilcoxon test

- Gives more weight to deaths at early time points
- The results can be misleading when a large fraction of patients is censored at early time points.
- Doesn't require a consistent hazard ratio, but does require that one group consistently have a higher risk than the other.

** Proportional hazards mean that the ratio of hazard functions (deaths per time) is the same at all time points.



30

NIAID

How to report statistical results

Overall:

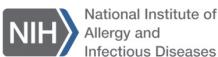
- Every statistical paper should report all methods completely enough so someone else could reproduce the work exactly.
- Every figure and table should present the data clearly
- All the results should be reported completely enough that no one wonders what you actually did.

Statistical Methods:

- State the full name of the test.
- Identify the program or the program that did the calculations
- State all options you selected. Repeated measures? Report enough detail so anyone could start with your data and get precisely the same results you got.

Graphing Data:

- Make main title, X- and Y-axes, legends, or any labels clear.
- Present data clearly. Focus on letting the reader see the data, and not only your conclusions.
- When possible, graph the individual data, not a summary of the data.



NIAID

31

Conclusions

- The materials of this seminar will be updated on: [Github - Graphing with Prism](#)
- If you have any further statistical problem, please send email to
bioinformatics@niaid.nih.gov
- Check our training schedule: http://www.eventzilla.net/user/NIAID_OCICB_BCBB



NIAID

32

Thank You!



National Institute of
Allergy and
Infectious Diseases

NIAID