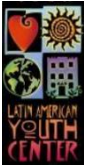


Dataviz with R and ggplot2



Tony Fujs

Latin American Youth Center

tony@layc-dc.org

@tonyfujs

The grammar of graphics:

In the grammar of graphics framework, a plot is made of the following elements:

	Elements of a plot	code
	coordinate system - <i>ONE per plot</i>	coord_xxxx
	scale - <i>ONE per aesthetic</i>	scale_xxxx
	facet – <i>ONE per plot (optional)</i>	
Layer At least one per plot, possibly many	<ul style="list-style-type: none">• data – <i>ONE per layer (can be the same)</i>• aesthetic mappings• geometric object – <i>ONE per layer</i>• position adjustment - <i>ONE per layer</i>• Statistical transformation - <i>ONE per layer</i>	<ul style="list-style-type: none">dataaesgeom_xxxxpositionstat

Cheat sheet:

For more information please refer to the ggplot2 help page: <http://docs.ggplot2.org/current/index.html>

- **Coordinate system:**

<i>coord_cartesian()</i>	The Cartesian coordinate system is the most familiar, and common, type of coordinate system.
<i>coord_polar()</i>	The polar coordinate system is most commonly used for pie charts, which are a stacked bar chart in polar coordinates.
<i>coord_flip()</i>	Flipped cartesian coordinates so that horizontal becomes vertical, and vertical, horizontal.
<i>coord_map()</i>	Map projection coordinate system

- **Scale:**

<i>scale_x_continuous()</i>	Continuous position scale for x
<i>scale_y_continuous()</i>	Continuous position scale for y
<i>scale_colour_continuous()</i>	Control the mapping of continuous variables to the color space
<i>scale_color_discrete()</i>	Control the mapping of discrete variables to the color space

- **Aesthetics:**

<i>x</i>	Map variable to geom position on the x axis
<i>y</i>	Map variable to geom position on the y axis
<i>size</i>	Map variable to size of geom
<i>Color</i>	Map variable to color of geom (exterior)
<i>Fill</i>	Map variable to color of geom (interior)
<i>Shape</i>	Map variable to shape of geom
<i>Alpha</i>	Map variable to transparency of geom

- **Geometric object:**

<i>geom_bar()</i>	The bar geom is used to produce 1d area plots: bar charts for categorical x, and histograms for continuous y.
<i>geom_point()</i>	The point geom is used to create scatterplots.
<i>geom_line()</i>	Connect observations, ordered by x value.
<i>geom_path()</i>	Connect observations in original order
<i>geom_polygon()</i>	Polygon, a filled path.
<i>geom_text()</i>	Textual annotations.
<i>geom_smooth()</i>	Add a smoothed conditional mean.

- **Position adjustment:**

<i>identity</i>	Don't adjust position
<i>dodge</i>	Adjust position by dodging overlaps to the side.
<i>Stack</i>	Stack overlapping objects on top of one another
<i>fill</i>	Stack overlapping objects on top of one another, and standardise to have equal height.

- **Statistical transformation:**

<i>identity</i>	No transformation
<i>smooth</i>	Aids the eye in seeing patterns in the presence of overplotting. Add new columns to dataset: predicted values from a regression, confidence intervals, etc.
<i>Stack</i>	Add a new column to dataset: # number of points in a specific bin

R resources: Where do I start?

1. Download and install R: <http://cran.r-project.org/>
2. Install R Studio (A user friendly interface for R): <http://www.rstudio.com/ide/download/>
3. Become familiar with the R environment: <http://tryr.codeschool.com/>
4. Learn how to install packages with R Studio: <http://www.youtube.com/watch?v=u1r5XTqrCTQ>
5. Take a free on-line class:
Beginner: <https://www.coursera.org/course/stats1>
Intermediate: <https://www.coursera.org/course/dataanalysis>
Advanced: <https://www.coursera.org/course/compdata>
6. Get started with your own "toy project" and stick to it!
There is a wealth of resources on-line and it's easy to get overwhelmed.
7. Minimize frustration by copy-pasting existing code for your "toy project"
Don't forget: "Google is your friend!"
8. Expect some frustration!

Additional resources:

- Download: <http://cran.r-project.org/doc/contrib/Short-refcard.pdf>
- Getting starting with R? Chances are what you are trying to do has already been documented <http://www.statmethods.net/> or <http://www.cookbook-r.com/>
- Watch Introduction to R, a video series by Google:
<http://www.youtube.com/playlist?list=PLOU2XLYxmsIK9qQfztXeybpHvru-TrqAP>
- 2 minutes video series: <http://www.twotorials.com/>

Ggplot2 resources:

Free on the web:

- R cookbook : <http://www.cookbook-r.com/Graphs/>
Good place to start. Many easy to follow examples with code.
- Official ggplot2 page: <http://docs.ggplot2.org/current/>
Most complete and up-to-date resource
- SAPE webpage: <http://sape.inf.usi.ch/quick-reference/ggplot2/>
Very complete and well organized resource

Books:

- ggplot2 Elegant Graphics for Data Analysis by Hadley Wickam
<http://www.amazon.com/ggplot2-Elegant-Graphics-Data-Analysis/dp/0387981403>

Other resources:

- The grammar of graphics by Leland Wilkinson
http://www.amazon.com/Grammar-Graphics-Statistics-Computing/dp/0387245448/ref=sr_1_1?s=books&ie=UTF8&qid=1398460257&sr=1-1