Instituto Tecnológico de Tijuana

Ingeniería en Sistemas Computacionales

Investigación #1

Materia: Minería de Datos

Unidad: Unidad II

Facilitador:

José Christian Romero Sánchez

Alumno: Hernández Negrete Juan Carlos - **16212021**

Fecha:

Tijuana Baja California a 04 de Mayo del 2021.

The grammar of graphics.

The grammar of the graphs move us from the paradigm of the types of graphs and focus on the elements that constitute a statistical graph, both those that are directly visible in the final result and those that, although they are not directly in view, configure the visible elements. By clearly identifying the elements we can chain and in the same way that we chain words to produce sentences we can chain elements to produce graphics and understand graphics.

Graph grammar elements			
Element	**Visible**		**Some...**
Geometric	Yes	They are the marks in space that represent the data.	Points, lines, polygons, rectangles
Data mapping	No	It is the specification of the data series that control properties of the geometric elements	x, y, color, fill, shape
Transformation	No	It is the specification of transformation of data mapped to representable data in space	Counts, means, standard errors, models
Scale	Yes	They indicate the relationship between a numerical series –variable– and the place that geometric elements occupy in space.	Axes, marks, labels
Coordinates	No	Control the projection of a number series to the two-dimensional space of a graph	Cartesian, polar, ternary, fixed
Annotations	Yes	They add information to the graph in the form of text, it is produced directly by the user	Title, subtitle, bullet.

From these elements we could define a bar chart as one whose geometric elements are rectangles, in which a categorical variable is mapped on the x-axis, which creates a continuous variable for the y-axis from a transformation that reports the counts for each category of that variable, which has a discrete scale for the x-axis and a continuous one on the y-axis and is represented in a Cartesian space.

The graph grammar applied in ggplot ().
In the grammar of graphs, they have five basic components from which we can control practically all aspects of a graph. ggplot () implements them in its syntax. Several of the elements are created by default, for example from the data mapping ggplot () creates the scale with its marks and labels. In the case of coordinates, it always takes a Cartesian system by default. In all cases we can subsequently modify these default values.

- A specification of mapping to variables.
- They can refer to both the axes of the graph (x, y)
- as well as other properties such as colors, symbols, sizes, fonts, line types, etc.
- In ggplot () they are specified with the arguments passed to aes ().
- A coordinate system.
- Controls the projection of numeric data to space.
- By default ggpplot () creates a Cartesian coordinate system with the x and y axes,
- We can change it to polar, ternary coordinates or map projections.
- It can be modified by adding + coord_ *, where * is any coordinate system included in ggplot () or its extensions.
- A scale.
- It controls the relationship between the number series that we are representing and its representation in space.
- By default ggplot () creates a natural scale for the axes in which they extend a little more than the maximum and minimum observations.
- By default ggplot () places the axes at 0 on the scale.
- We can change the scale through a transformation, for example, going from natural to logarithmic scale.
- By default ggplot () creates scale marks and labels in closed numbers (units, tens, hundreds, thousands, etc.), depending on the range of our data.
- It is possible to change the number and location of scale marks and / or labels.
- It can be modified by adding + scale _ * _ °, where * is the axis that we are going to modify and ° the type of scale that we are using: continuous or discrete.
- A statistical transformation.
- It controls the necessary transformation of the data for its representation.
- For example, you perform frequency counts for a bar graph or histogram.
- Or more complex transformations like those necessary for a boxplot.
- It does not apply to all graphs and there is the possibility of doing it ourselves and passing the transformed data to ggplot ().
- It can be modified with the argument stat = "*", where * is the transformation we want to perform. stat = identity to not do any transformation and we use it when provided by the user.
- A geometric element.
- It is the element that represents the data on the chart.
- We can use different elements to represent the same data.
- Or combine several in the same graph.
- Full labels.
- ggplot () takes variable names for axes and labels for scales (discrete or continuous)
- Of the additional mappings (that is, those that do not refer directly to the axes) the legend labels.
- Annotations.

What is ggplot2?
It is an R package specifically designed to produce graphics, but unlike other packages, ggplot2 has its own grammar or programming language. This grammar is based on the "Graphical Grammar" (Wilkinson 2005), which is made up of independent components that can be composed in many ways. This allows ggplot2 to create very flexible graphs, specific to any situation or problem. The ggplot2 package was born in 2005, with the idea of taking and improving the good of the “base” and “lattice” graphing that existed in R, to create a more robust graphing model.

In ggplot2 we can start with a raw data layer and then add more layers of annotations and statistical summaries. This package allows us to produce graphs using the same thinking structure that we use when designing an analysis, reducing the distance of how we visualize a graph in the head and the final product in our worksheet.

Learning the grammar for creating graphs in ggplot2 will not only be crucial for producing an interesting graph, but also for thinking about other more complex graphs. In R base charts, when we design a chart, it is typically made up of raw graphing elements such as points and lines, and it is difficult to design new components that blend in with existing charts. In ggplot2, the expressions used to create a new graph are made up of more complex elements from the raw data, which allow them to be easily combined with other data sets or graphs.

What is the grammar of ggplot2?
Wilkinson (2005) was responsible for creating the graphing grammar to describe all the attributes that are part of statistical graphs. The book "The layer grammar of graphics" by Wickham (2010) builds on Wilkinson's grammar, focusing on the importance of layers and adapting them to R. In general, the grammar of ggplot2 tells us that a statistical graph is a mapping of data to aesthetic attributes (color, shape, size) and geometric objects (points, lines, bars). The graph can also contain statistical transformations of the data and is drawn in a specific coordinate system.

Components (edit)
All charts are composed of:

Data - the data we want to visualize and a set of aesthetic elements that describe how the variables in the data will be mapped to aesthetic attributes that we can perceive.
Layers, which are made up of geometric elements (geom points, lines, polygons, etc.) and statistical or stat transformations (for example, collecting and counting observations to produce histograms or summarize relationships from a linear model).
scale - the scale, maps values in data space to values in aesthetic space, be it color, size or shape. The scale draws the legend or axes (X-Y).
coord. - the coordinate system describes how the data coordinates are mapped on the data plane. It also produces the axes and grid lines so that the graph can be read better. Normally a Cartesian coordinate system is used, but there are other systems available (polar coordinates, for example).

facet - describes how to break data into subsets and how to display these subsets in smaller multiples.
theme - controls the finer elements like font size and background color.

Other graphics packages

Currently there are other packages that have been created to solve specific situations, using a bit of the ggplot2 graphing model as a base. Among them the following can be mentioned:

- ggvis
- vcd (Warnes 2015)
- plotrix (Lemon et al. 2006)
- gplots (Warnes 2015)

REFERENCES

- Paladino. (2020). Graficos en R con ggplot. 04 de Mayo del 2021, de Instituto Mora Sitio web: https://www.institutomora.edu.mx/testU/SitePages/martinpaladino/graficos_con_r_y_ggplot.html

- Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2015) gplots: various R programming tools for plotting data. R package version 2.17.0. https://CRAN.R-project.org/package=gplots