An Introduction to ggvis

Anjali Nadeswaran

11/28/2017

Introduction

ggivis is a data visualization package for R that is used to create interactive graphs for exploratory data analysis.

The syntax and functions of ggvis are very similar to those of the package ggplot2. However, ggvis is specifically important because of its integration with shiny's reactive programming model and dplyr's grammar of data transformation.

The strengths of ggvis are in its features that build interactive plots. Interactive abilities make the package useful for exploration purposes, but harder to use in publications. This is because interactive ggvis plots must be connected to a running R session to be viewed.

Background

The package ggvis was created by Winston Chang and Hadley Wickham in order to take the best parts of ggplot2 and incorporate them with the reactive framework of shiny and drawing web graphics using vega.

Applications

With ggvis, we can build graphics with multiple layers.
The five simple layers of the package are:
-points; layer_points( )
-paths and polygons; layer_paths( )
-filled areas; layer_ribbons( )
-rectangles; layer_rects( )
-text; layer_text( )

The five most common compound layers include:
-lines; layer_lines( )
-histograms; layer_histograms( )
-predictive models; layer_smooths( )

We can then connect these visual properties to interactive controls that allow the user to control the size of points, the width of histogram bins, etc. In this tutorial, we will explore the more complex interactive feature of ggvis: tooltips.

Tooltips

The function add_tooltip allows us to program certain behavior when we hover or click on a point. We can provide a function that treats the data stored in a given point as an input.

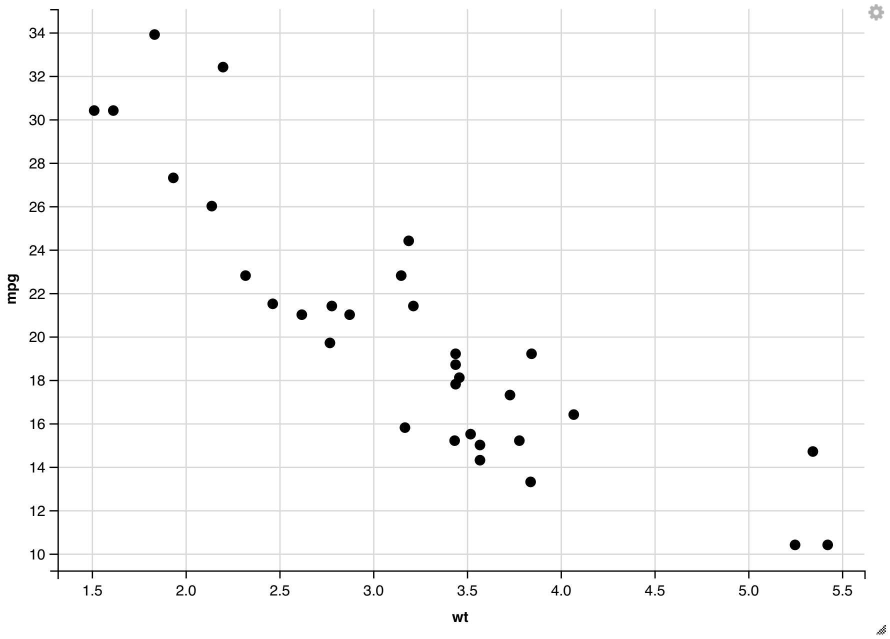
Examples

Let's begin by creating a static plot with ggvis. For this example, we will work with the dataset, mtcars, that is built into R. The data was extracted from the 1974 *Motor Trend* US magazine. It consists of a data frame with 32 observations on 11 variables.

```
#load packages ggvis and shiny
library(ggvis)
library(shiny)

#add column id to dataframe, labeling the rows from 1 to 32
mtcars$id <- 1:nrow(mtcars)

mtcars %>% ggvis(x = ~wt, y = ~mpg, key := ~id) %>%
  layer_points()
```



To add another layer to the graph, we can add a tooltip. In this example, we will use add_tooltip so that when we hover over a point on the graph it displays the point's values for each of the 12 variables in the dataset.

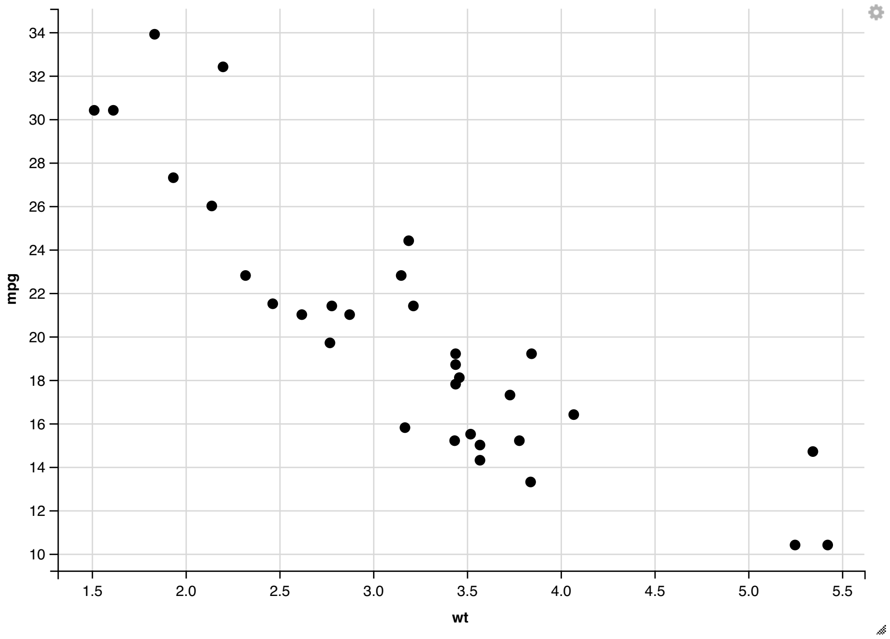
In order to do so, we must first create a single function that prints each observation's data for all variables in mtcars.

```
#create function all_values that prints data in each row
all_values <- function(x) {
  row <- mtcars[mtcars$id == x$id, ]
  paste0(names(row), ":", format(row), collapse = "<br />") #prints values in a cleaner format
}
```

Now, we can feed this new function through add_tooltip to complete our graph.

```
mtcars %>% ggvis(x = ~wt, y = ~mpg, key := ~id) %>%
  layer_points() %>%
  add_tooltip(all_values, "hover")
```

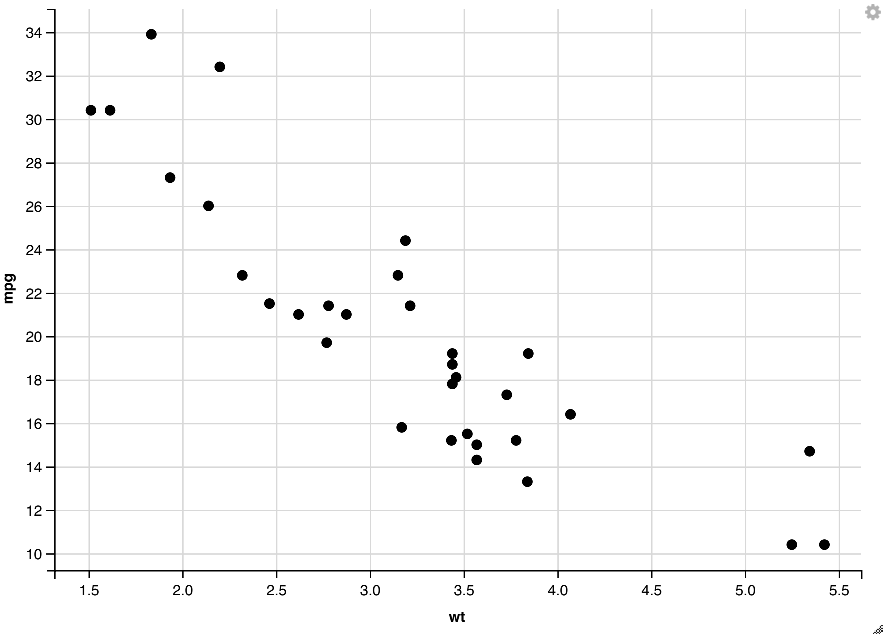
```
## Warning: Can't output dynamic/interactive ggvis plots in a knitr document.
## Generating a static (non-dynamic, non-interactive) version of the plot.
```



We can also alter this tooltip so that it shows the data associated with each point when we click on the point, instead of hover on it.

```
mtcars %>% ggvis(x = ~wt, y = ~mpg, key := ~id) %>%
  layer_points() %>%
  add_tooltip(all_values, "click")
```

```
## Warning: Can't output dynamic/interactive ggvis plots in a knitr document.
## Generating a static (non-dynamic, non-interactive) version of the plot.
```



Keep in mind that interactive graphs created with ggvis must be viewed in a running R session. So, in this knitted html file, we will only see a static version of the plot that matches the graph we created earlier.

Discussion

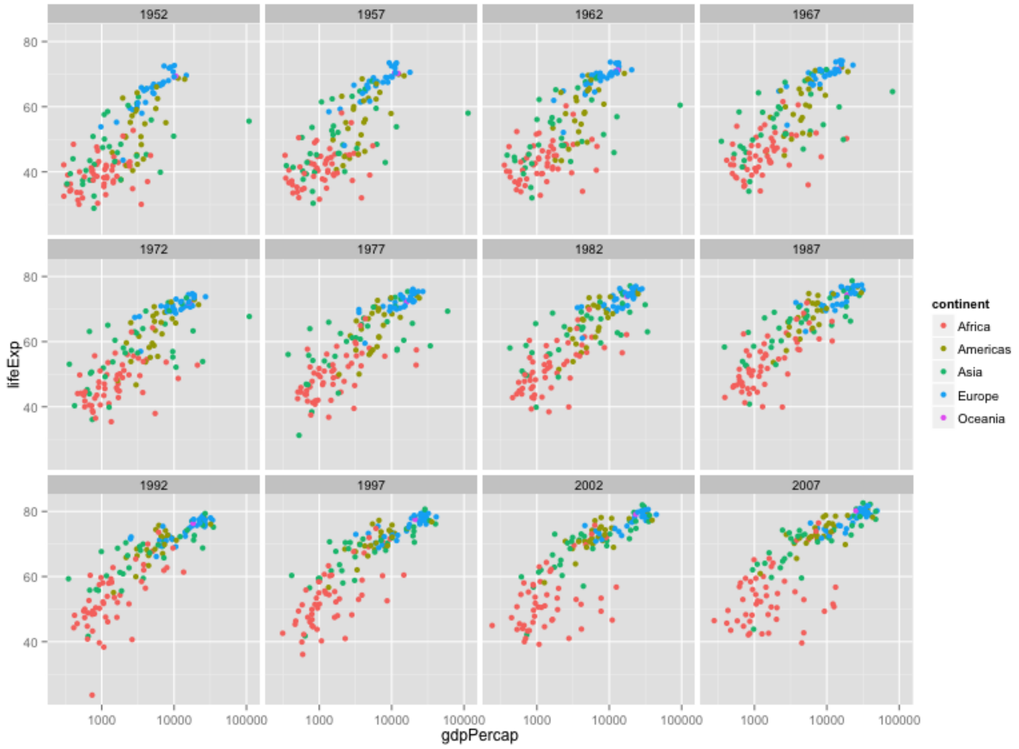
ggvis vs ggplot2

From our example, we can see that the syntax we used to create the ggvis plot is very similar to the syntax used to create graphs with ggplot2.

In effect, the function ggplot( ) in the ggplot2 package is equivalent to the function ggvis( ) in the ggvis package. However, one important difference between the two functions is the use of piping (%>%) in ggvis where ggplot2 uses "+".

When to Use ggvis

Because ggplot2 is a more mature package, it is more stable and less likely to procure bugs. For this reason, it is usually the preferred tool to create plots if we are not trying to make an interactive graph. ggvis also lacks faceting, a very essential part of ggplot2. Faceting allows us to plot relationships between variables in multiple subsets of the data, visualized as panels in a larger figure. Here is an example of this feature, below.



Faceting in ggplot2

Due to these differences, ggplot2 is primarily used for printed publications, while ggvis is used for interactive plots in reports on data. For the same reasons, it is important to identify the end goal of your research and know when to use one package over another to fit your specific purposes.

Conclusion

From the applications of ggvis that we discussed, we can see that it is an ideal tool for exploratory data analysis. Exploratory data analysis postpones the typical assumptions about the type of model that the data follows through the more direct approach of allowing the data itself to reveal its underlying structure and model. This approach to data analysis is usually graphical in nature as graphs allow us to explore data with an open mind. The interactive capabilities of ggvis graphs are especially useful because they allow data analysts to play with data, visualize effects, and test underlying assumptions.

The function add_tooltip can be used to add another advantageous interactive feature to ggvis plots. By displaying certain data about a point when we hover or click on it, we can create an informative graph that is still clean and easy to read. For this purpose, it is a very important function to learn when using ggvis for exploratory data analysis.

Resources

https://ggvis.rstudio.com/
http://www.londonr.org/presentations/2014/11/LondonR_-_Introduction_to_ggvis_-_Aimee_Gott_-_20141125.pdf
https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html
https://cran.r-project.org/web/packages/ggvis/ggvis.pdf
http://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm
https://www.rstudio.com/resources/webinars/the-grammar-and-graphics-of-data-science/
https://ggvis.rstudio.com/ggplot2.html
https://www3.nd.edu/~steve/computing_with_data/13_Facets/facets.html