

# Infographics in R

Anna Miera

December 3rd, 2017

## Introduction

I am interested in the interaction between nonprofit organizations and data analysts. Previously, I explored data analysis methods for qualitative and quantitative data. Now I want to see what R can do to make data visuals more eye-catching and friendly by researching ways to make infographics and pictograms in R.

For an example, we will say that the nonprofit that we work for is called **Better Beginnings** (BB) and they support families with infants. Their program works one-on-one with families, encouraging them to use developmentally appropriate practices. We want to make them data visuals that show:

- Growth on a “baby development test” over time for control and experimental group
- Number of books read in control and experimental group

## Reproducible Code

Here is the data that I created in order to make the code reproducible along with the R packages that I used. Feel free to copy and paste onto your console.

### Data

The nonprofit has information about a control group and families that went through their program. Their program lasted 6 months and a baby development test was administered when the baby was 3, 6, and 9 months. The number of books read by the parents to their child in the 6 months was also recorded.

```
# vector of family last names
name <- c(LETTERS[1:12])

# vector of control vs Better Beginnings program
group <- c(rep("BB", 6), rep("control", 6))

# vector of baby development test results (say the score can be
# from 0 to 100)
month3 <- c(33, 52, 25, 40, 10, 15, 32, 17, 58, 24, 42, 29)
month6 <- c(40, 57, 38, 51, 32, 38, 42, 26, 63, 31, 46, 35)
month9 <- c(59, 68, 57, 69, 46, 46, 53, 38, 71, 45, 53, 45)

# vector of number of books read in 6 months
books <- c(101, 77, 206, 124, 168, 111, 66, 25, 146, 31, 68, 25)

# data frame with all info
dat <- data.frame(name, group, month3, month6, month9, books)
```

### Packages

Packages may need to be installed on your computer. Use `install.packages("")` to have these accessible.

```
# packages used under R version 3.4.3
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(png)
```

```
## Loading required package: png
```

```
require(grid)
```

```
## Loading required package: grid
```

```
require(waffle)
```

```
## Loading required package: waffle
```

```
## Warning: package 'waffle' was built under R version 3.4.3
```

## Scatterplot of Books Read vs Test Growth

Now let's create some visuals for the nonprofit! We will first make the scatterplot.

Clara Schartner (data scientist) writes a post called, [“Putting the cat in scatterplot,”](#) that uses ggplot to literally show pictures of cats on her plot. They helped me outline how to form a basic plot with images.

## Blank Plot

First, we will add a *progress* column to our data frame that will show the difference between scores of each baby from 3 months to 9 months on the child development test. Then, as Scharfner details, we need to make a blank ggplot by making the size of the points be zero. We will plot the number of books read with the growth on the test from 3 months to 9 months while considering whether the family was in the control group or the group that went through BB's program.

```
# adding a progress column
dat$progress <- (dat$month9 - dat$month3)

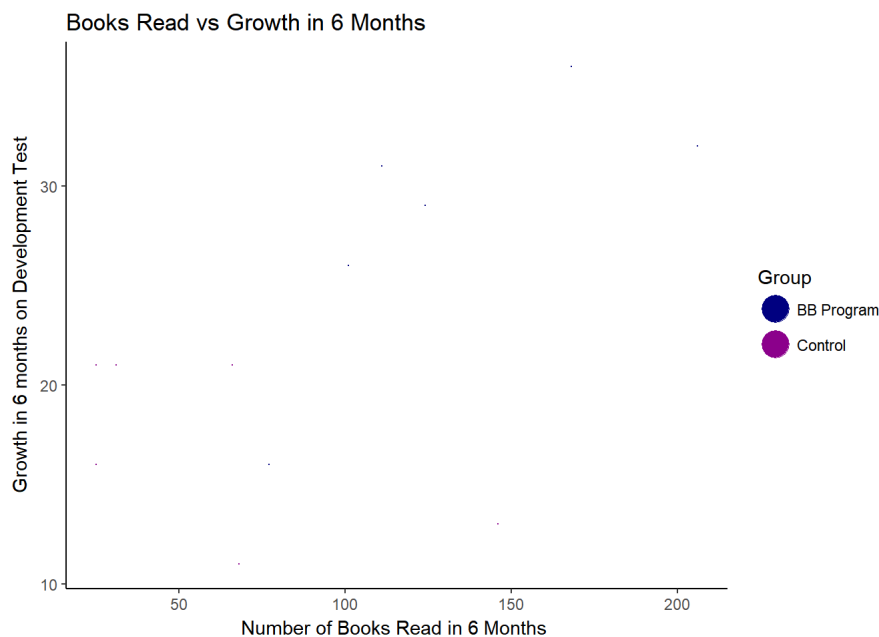
# ggplot base
growth <- ggplot(dat, aes(books, progress)) +

  # size zero so that the pictures will cover the point
  geom_point(size = 0, aes(group = group, color = group)) +
  labs(x = "Number of Books Read in 6 Months ",
       y = "Growth in 6 months on Development Test",
       title = "Books Read vs Growth in 6 Months") +

  # so the background is blank
  theme_classic() +

  # customized legend
  scale_color_manual(
    # our choice of colors for the labeling
    values = c("navy blue", "darkmagenta"),
    name = "Group",
    labels = c("BB Program", "Control")) +
  # size of the key
  guides(color = guide_legend(override.aes = list(size = 7)))

# display our blank plot
growth
```



## Adding the Images

Now we want to find an image that will represent our data but also work nicely with R. We can save a png image with the readPNG function in the png package. A nice and simple explanation about this technique can be found on [this website](#).

I found pictures (pngs) of the books I wanted to use to represent the data. A google image search works well but keep in mind that images with transparent backgrounds are recommended (typically clipart with a checkered background is good). Also, later we will want to overlap these books so a transparent background is needed in this case. The blue book came from [here](#) and the purple from [here](#). You should save the images as png files from the links above. Remember what you name them and save them in the same folder as your console's working directory (or another location that you know you can access).

I called the blue book, *bookb.png*, and the purple, *bookp.png*, and saved them in a *data* folder within my post's folder.

```
# saving the images
# blue book
bookb <- readPNG("../data/bookb.png")

# purple book
bookp <- readPNG("../data/bookp.png")
```

Now the books will be plotted using a for loop, "annotation\_custom()" (from ggplot), and *rasterGrob()* (from the package: grid). Norbert Köhler writes about these functions in this [article](#). He demonstrates how to use the functions to change the background in ggplot graphs.

```
# plotting the books
for (i in 1:nrow(dat)) {

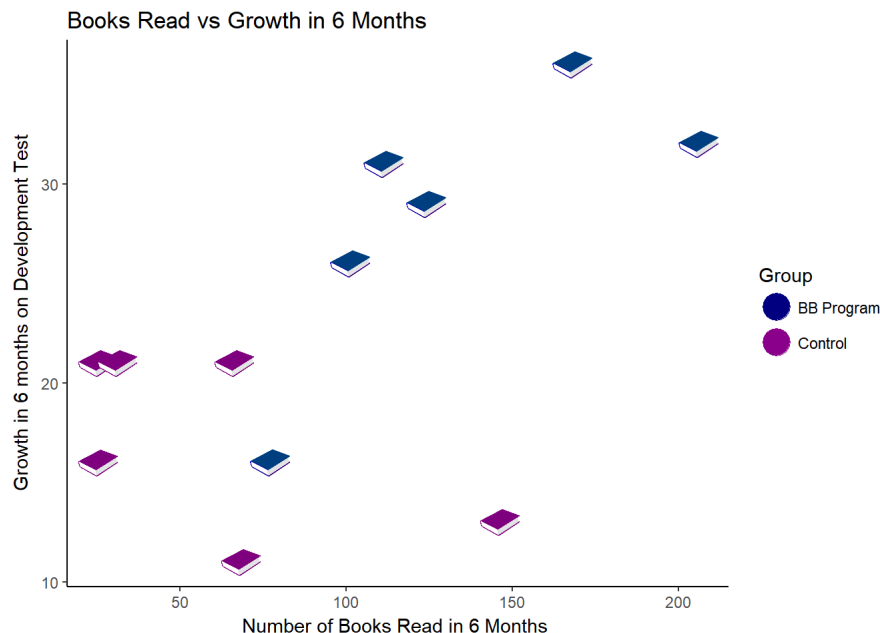
  # sorting by the group that the family is a part of
  if (dat$group[i] == "BB") {
    pic <- bookb
  } else{
    pic <- bookp
  }

  # changing the ggplot growth
  growth = growth +
  annotation_custom(

    # coerces image to a raster object
    rasterGrob(pic),

    # size of the books and location on graph
    xmin = dat$books[i] - 6,
    xmax = dat$books[i] + 6,
    ymin = dat$progress[i] - 6,
    ymax = dat$progress[i] + 6
  )
}

# display
growth
```



I would recommend trying to find the same color as your picture by seeing what [colors](#) R has available.

## Linear Regression Line

Now we can add a linear regression line that also plots a picture. Say BB wants us to use brains that will be plotted in a line to represent the linear regression. Again, I found the brain I wanted to use with Google and [this](#) is the one I picked. I saved it as *brain.png*.

```
# saving the brain image
brain <- readPNG("../data/brain.png")
```

We will perform linear regression using the *lm()* and *predict()* functions.

```
# linear model to carry out regression
lm <- lm(progress~books, data = dat)

# the number of brains that will be plotted to represent the linear regression
num_of_brains <- 20

# making a data frame with points
lr <- data.frame(books =
  seq(min(dat$books), max(dat$books),
    # number of predictions that will be made
    length.out = num_of_brains))

# using linear regression, predict values for progress
lr$progress <- predict(lm, newdata = lr)

# display what we found
lr
```

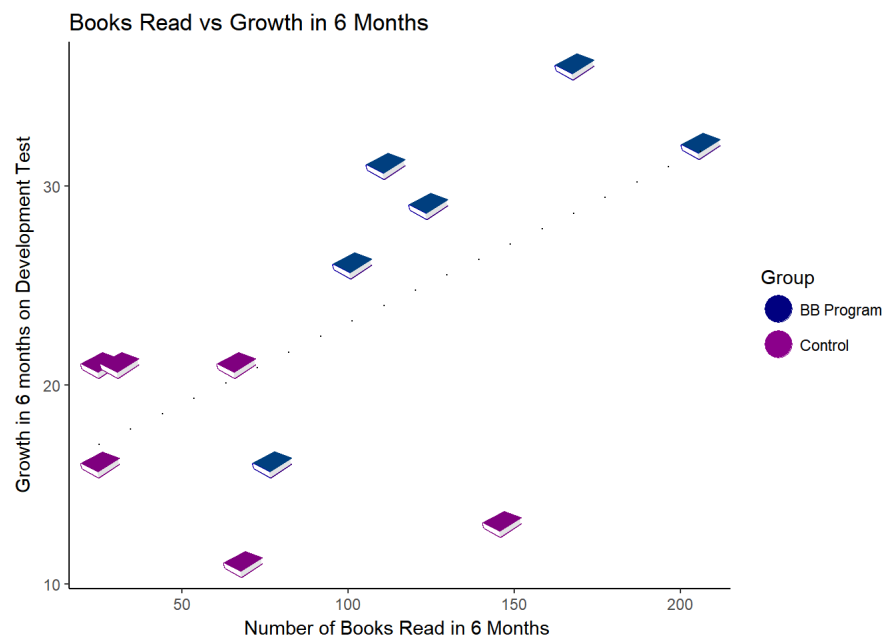
```
##      books progress
## 1  25.00000 16.99684
## 2  34.52632 17.77240
## 3  44.05263 18.54796
## 4  53.57895 19.32353
## 5  63.10526 20.09909
## 6  72.63158 20.87465
## 7  82.15789 21.65021
## 8  91.68421 22.42578
## 9 101.21053 23.20134
## 10 110.73684 23.97690
## 11 120.26316 24.75247
## 12 129.78947 25.52803
## 13 139.31579 26.30359
## 14 148.84211 27.07915
## 15 158.36842 27.85472
## 16 167.89474 28.63028
## 17 177.42105 29.40584
## 18 186.94737 30.18140
## 19 196.47368 30.95697
## 20 206.00000 31.73253
```

Everything now is ready and we just need to use the previous method to plot the brains along the regression line.

```
# adding the linear regression points of size 0 to our plot
growth <- growth +

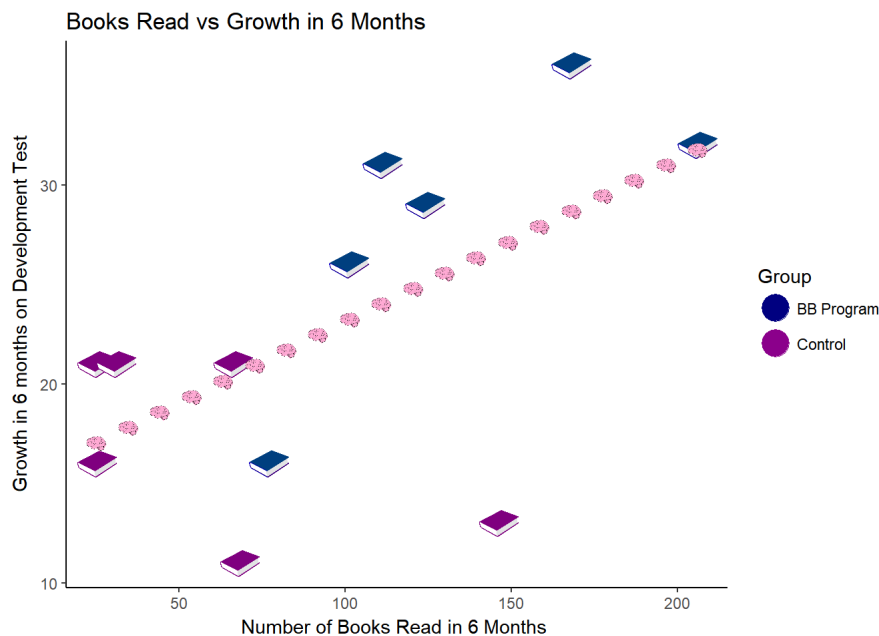
  geom_point(data = lr, aes(books, progress), size = 0)

# display the blank plot
growth
```



```
# using a for loop to plot the brains
for (i in 1:20){
  growth = growth +
    annotation_custom(
      rasterGrob(brain),
      xmin = lr$books[i] - 3,
      xmax = lr$books[i] + 3,
      ymin = lr$progress[i] - 3,
      ymax = lr$progress[i] + 3
    )
}

# display our final plot
growth
```



## Number of Books Read in Control vs Experimental

The nonprofit will want to showcase its most impressive results such as how many more books were read in the group that participated in BB's program. A pictorial barchart can offer a more friendly representation.

```
# number of books read by each group
# BB group
books_BB <- sum(dat$books[1:6])

# control group
books_control <- sum(dat$books[7:12])

# data frame of this data
books_data <- data.frame("num_read" = c(books_control, books_BB),
                        "group" = c("Control", "BB Program"))

# display data
books_data
```

```
##   num_read   group
## 1     361  Control
## 2     787 BB Program
```

## Blank Plot

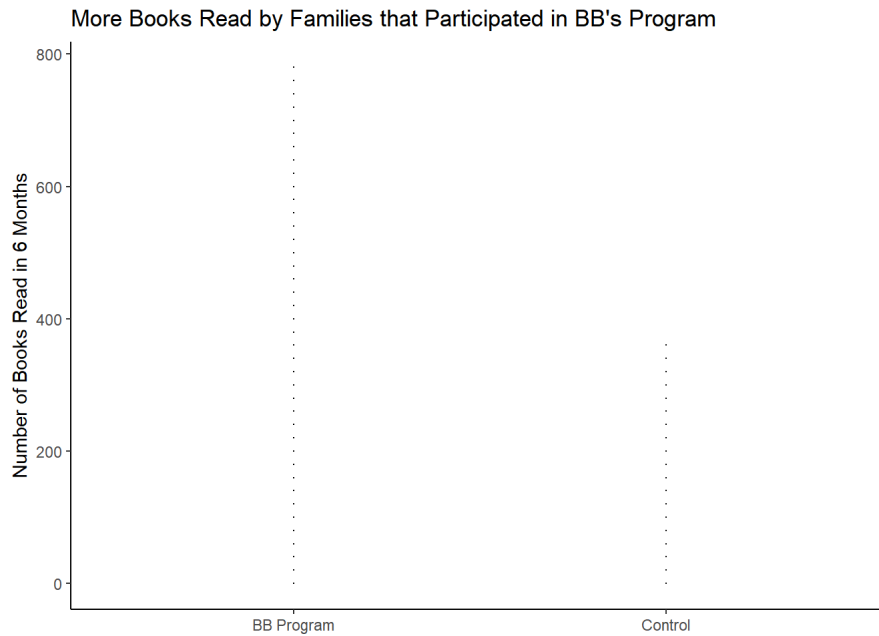
I searched for similar graphs made in R and found it helpful to review the STHDA's (Statistical Tools for High-throughput Data Analysis) [guide](#) to customizing labels in ggplot. We will use the same approach that was used to create the scatterplot by starting with a blank plot; More points will be added and will each be covered by a books so that a "barchart" of books will be displayed.

```
# creating vectors to represent the multiple points
# each image of a book will represent 20 books read
# NOTE: I rounded the number of books read in each group
num_read <- c(seq(0, 780, 20), seq(0, 360, 20))
group <- c(rep("BB Program", (780 / 20) + 1), rep("Control", (360 / 20) + 1))

# creating the data frame to represent this data
book_barchart <- data.frame(num_read, group)

# blank ggplot
book_stack <- ggplot(book_barchart, aes(group, num_read)) +
  geom_point(size = 0) +
  theme_classic() +
  theme(axis.title.x = element_blank()) +
  labs(y = "Number of Books Read in 6 Months",
       title = "More Books Read by Families that Participated in BB's Program")

# display blank barchart
book_stack
```



## Adding the Book Stacks

`bookb` and `bookp` will be used again to represent each group. But before this, our plot should have values for the x-axis or using annotation custom will be tricky (I tried).

```
# creating vectors for the x-axis so that we can use annotation custom
# BB Program = 1 and Control = 0
x_axis <- c(rep(1, (780 / 20) + 1), rep(0, (360 / 20) + 1))

# updating our data frame
book_barchart$x <- x_axis

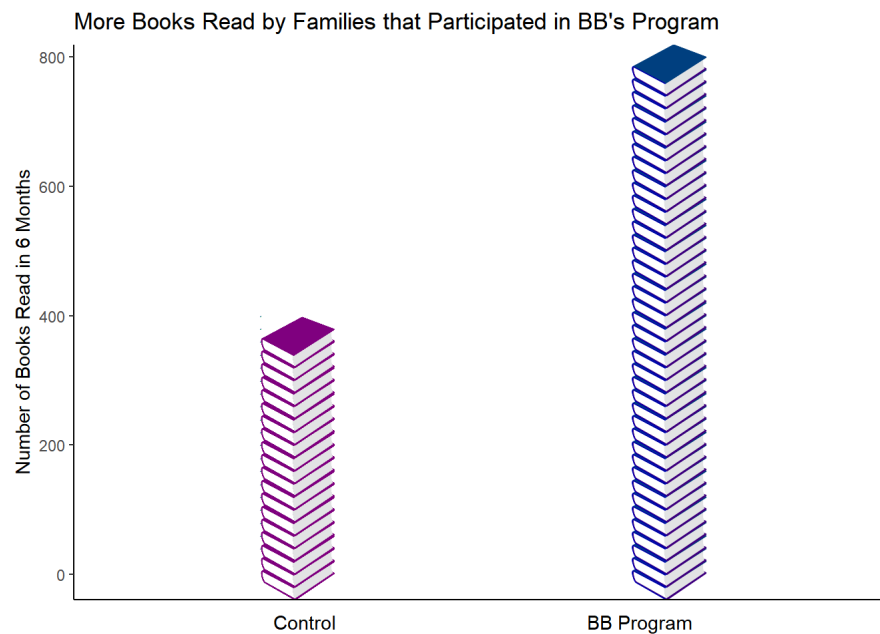
# updating our plot
book_stack <- ggplot(book_barchart, aes(x, num_read)) +
  geom_point(size = 0) +
  theme_classic() +
  # removing labels since the x value is unimportant
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank()) +
  # setting the x limits so the spacing looks normal
  xlim(-.5, 1.5) +
  labs(x = "Control", "BB Program",
       y = "Number of Books Read in 6 Months",
       title = "More Books Read by Families that Participated in BB's Program")

# using a for loop to see which color book should be used
for (i in 1:nrow(book_barchart)) {

  # sorting by the group that the family is a part of
  if (book_barchart$group[i] == "BB Program") {
    pic <- bookb
  } else{
    pic <- bookp
  }

  # updating the plot book_stack
  book_stack = book_stack +
    annotation_custom(
      rasterGrob(pic),
      xmin = book_barchart$x[i] - 20,
      xmax = book_barchart$x[i] + 20,
      ymin = book_barchart$num_read[i] - 40,
      ymax = book_barchart$num_read[i] + 40
    )
}

# display our final plot
book_stack
```



## Conclusions

Okay so making infographics in R is a lot trickier than I thought. And although it allows you to customize more than other plot-making online tools, for simple plots it can really overcomplicate things. [This article](#) shows how to add your plots to an [infographic-making tool](#) to make your final result looks pretty. But if you really are working for a nonprofit organization as their graphic designer, then you should probably learn how to make graphic designs.

Anyways, I hope that you enjoyed this *very basic* introduction to infographics in R!

## References

- **Putting the cat in scatterplot:** <https://www.mango-solutions.com/blog/putting-the-cat-in-scatterplot>
- **ReadPNG:** <http://rfunction.com/archives/2539>
- **ggplot Backgrounds:** <https://datascienceplus.com/how-to-add-a-background-image-to-ggplot2-graphs/>
- **R's Colors:** <http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf>
- **STHDA's Labeling Guide:** <http://www.sthda.com/english/wiki/ggplot2-axis-ticks-a-guide-to-customize-tick-marks-and-labels>
- **Infographics in R:** <http://nandeshwar.info/data-visualization/how-to-create-infographics-in-r/>
- **Infographic Online Tools:** <http://www.creativebloq.com/infographic/tools-2131971>