

Data Visualization in R

Vanilla R vs. ggplot2

- Vanilla R or base R – a blank canvas you build on
- ggplot2 – set of tools to use with constraints

Vanilla R vs. ggplot2



"good enough" is usually below the asymptote, but not always.



4:58 PM · May 27, 2021 · Twitter for iPhone

Making Figures with Vanilla R

Open and closing an output PDF file

```
> pdf.options(family="Helvetica")
> pdf("output_file.pdf", pointsize=10)
...
> dev.off()
```

NOTE: > to indicates an R command

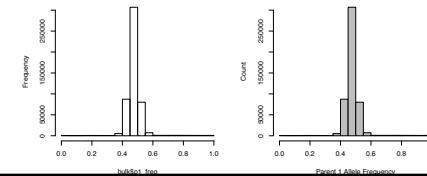
* It can be easy to forget dev.off(). If you are working in RStudio and you can't produce chart it may be because you forgot the dev.off() command after using pdf().

Making a scatter plot between two variables

```
> plot(var$col_1, var$col_2)
> plot(var$col_1, var$col_2, xlab="Column 1", ylab="Column 2")
```

Making a histogram plot

```
> hist(var$col_1)
> hist(var$col_1, xlab="P1 Allele Frequency", ylab="Count", col="gray")
```



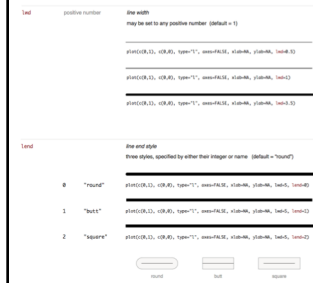
Other Features That Can Be Helpful

Adding lines to a plot

```
> lines(c(2,7),c(0, 500), col="red", lty=1, lwd=2)
```

Adding boxes to a plot

```
> rect(455861, -0.1, 13854546, 1.1, col="#99330025", border=NA)
```



R_lines.pdf

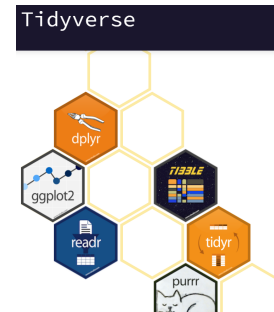
color	name	color	name
	white	black	black0a
	black		black0a
	gray0a		black0a1
	gray0a1		black0a2
	gray0a2		black0a3
	gray0a3		black0a4
	gray0a4		black0a5
	gray0a5		black0a6
	gray0a6		black0a7
	gray0a7		black0a8
	gray0a8		black0a9
	gray0a9		black0aa
	gray0aa		black0ab
	gray0ab		black0ac
	gray0ac		black0ad
	gray0ad		black0ae
	gray0ae		black0af
	gray0af		black0b0
	gray0b0		black0b1
	gray0b1		black0b2
	gray0b2		black0b3
	gray0b3		black0b4
	gray0b4		black0b5
	gray0b5		black0b6
	gray0b6		black0b7
	gray0b7		black0b8
	gray0b8		black0b9
	gray0b9		black0ba
	gray0ba		black0bb
	gray0bb		black0bc
	gray0bc		black0bd
	gray0bd		black0be
	gray0be		black0bf
	gray0bf		black0c0
	gray0c0		black0c1
	gray0c1		black0c2
	gray0c2		black0c3
	gray0c3		black0c4
	gray0c4		black0c5
	gray0c5		black0c6
	gray0c6		black0c7
	gray0c7		black0c8
	gray0c8		black0c9
	gray0c9		black0ca
	gray0ca		black0cb
	gray0cb		black0cc
	gray0cc		black0cd
	gray0cd		black0ce
	gray0ce		black0cf
	gray0cf		black0d0
	gray0d0		black0d1
	gray0d1		black0d2
	gray0d2		black0d3
	gray0d3		black0d4
	gray0d4		black0d5
	gray0d5		black0d6
	gray0d6		black0d7
	gray0d7		black0d8
	gray0d8		black0d9
	gray0d9		black0da
	gray0da		black0db
	gray0db		black0dc
	gray0dc		black0dd
	gray0dd		black0de
	gray0de		black0df
	gray0df		black0e0
	gray0e0		black0e1
	gray0e1		black0e2
	gray0e2		black0e3
	gray0e3		black0e4
	gray0e4		black0e5
	gray0e5		black0e6
	gray0e6		black0e7
	gray0e7		black0e8
	gray0e8		black0e9
	gray0e9		black0ea
	gray0ea		black0eb
	gray0eb		black0ec
	gray0ec		black0ed
	gray0ed		black0ee
	gray0ee		black0ef
	gray0ef		black0f0
	gray0f0		black0f1
	gray0f1		black0f2
	gray0f2		black0f3
	gray0f3		black0f4
	gray0f4		black0f5
	gray0f5		black0f6
	gray0f6		black0f7
	gray0f7		black0f8
	gray0f8		black0f9
	gray0f9		black0fa
	gray0fa		black0fb
	gray0fb		black0fc
	gray0fc		black0fd
	gray0fd		black0fe
	gray0fe		black0ff
	gray0ff		black100
	gray100		black101
	gray101		black102
	gray102		black103
	gray103		black104
	gray104		black105
	gray105		black106
	gray106		black107
	gray107		black108
	gray108		black109
	gray109		black10a
	gray10a		black10b
	gray10b		black10c
	gray10c		black10d
	gray10d		black10e
	gray10e		black10f
	gray10f		black110
	gray110		black111
	gray111		black112
	gray112		black113
	gray113		black114
	gray114		black115
	gray115		black116
	gray116		black117
	gray117		black118
	gray118		black119
	gray119		black11a
	gray11a		black11b
	gray11b		black11c
	gray11c		black11d
	gray11d		black11e
	gray11e		black11f
	gray11f		black120
	gray120		black121
	gray121		black122
	gray122		black123
	gray123		black124
	gray124		black125
	gray125		black126
	gray126		black127
	gray127		black128
	gray128		black129
	gray129		black12a
	gray12a		black12b
	gray12b		black12c
	gray12c		black12d
	gray12d		black12e
	gray12e		black12f
	gray12f		black130
	gray130		black131
	gray131		black132
	gray132		black133
	gray133		black134
	gray134		black135
	gray135		black136
	gray136		black137
	gray137		black138
	gray138		black139
	gray139		black13a
	gray13a		black13b
	gray13b		black13c
	gray13c		black13d
	gray13d		black13e
	gray13e		black13f
	gray13f		black140
	gray140		black141
	gray141		black142
	gray142		black143
	gray143		black144
	gray144		black145
	gray145		black146
	gray146		black147
	gray147		black148
	gray148		black149
	gray149		black14a
	gray14a		black14b
	gray14b		black14c
	gray14c		black14d
	gray14d		black14e
	gray14e		black14f
	gray14f		black150
	gray150		black151
	gray151		black152
	gray152		black153
	gray153		black154
	gray154		black155
	gray155		black156
	gray156		black157
	gray157		black158
	gray158		black159
	gray159		black15a
	gray15a		black15b
	gray15b		black15c
	gray15c		black15d
	gray15d		black15e
	gray15e		black15f
	gray15f		black160
	gray160		black161
	gray161		black162
	gray162		black163
	gray163		black164
	gray164		black165
	gray165		black166
	gray166		black167
	gray167		black168
	gray168		black169
	gray169		black16a
	gray16a		black16b
	gray16b		black16c
	gray16c		black16d
	gray16d		black16e
	gray16e		black16f
	gray16f		black170
	gray170		black171
	gray171		black172
	gray172		black173
	gray173		black174
	gray174		black175
	gray175		black176
	gray176		black177
	gray177		black178
	gray178		black179
	gray179		black17a
	gray17a		black17b
	gray17b		black17c
	gray17c		black17d
	gray17d		black17e
	gray17e		black17f
	gray17f		black180
	gray180		black181
	gray181		black182
	gray182		black183
	gray183		black184
	gray184		black185
	gray185		black186
	gray186		black187
	gray187		black188
	gray188		black189
	gray189		black18a
	gray18a		black18b
	gray18b		black18c
	gray18c		black18d
	gray18d		black18e
	gray18e		black18f
	gray18f		black190
	gray190		black191
	gray191		black192
	gray192		black193
	gray193		black194
	gray194		black195
	gray195		black196
	gray196		black197
	gray197		black198
	gray198		black199
	gray199		black19a
	gray19a		black19b
	gray19b		black19c
	gray19c		black19d
	gray19d		black19e
	gray19e		black19f
	gray19f		black200
	gray200		black201
	gray201		black202
	gray202		black203
	gray203		black204
	gray204		black205
	gray205		black206
	gray206		black207
	gray207		black208
	gray208		black209
	gray209		black20a
	gray20a		black20b
	gray20b		black20c
	gray20c		black20d
	gray20d		black20e
	gray20e		black20f
	gray20f		black210
	gray210		black211
	gray211		black212
	gray212		black213
	gray213		black214
	gray214		black215
	gray215		black216
	gray216		black217
	gray217		black218
	gray218		black219
	gray219		black21a
	gray21a		black21b
	gray21b		black21c
	gray21c		black21d
	gray21d		black21e
	gray21e		black21f
	gray21f		black220
	gray220		black221
	gray221		black222
	gray222		black223
	gray223		black224
	gray224		black225
	gray225		black226
	gray226		black227
	gray227		black228
	gray228		black229
	gray229		black22a
	gray22a		black22b
	gray22b		black22c
	gray22c		black22d
	gray22d		black22e
	gray22e		black22f
	gray22f		black230
	gray230		black231
	gray231		black232
	gray232		black233
	gray233		black234
	gray234		black235
	gray235		black236
	gray236		black237
	gray237		black238
	gray238		black239
	gray239		black23a
	gray23a		black23b
	gray23b		black23c
	gray23c		black23d
	gray23d		black23e
	gray23e		black23f
	gray23f		black240
	gray240		black241
	gray241		black242
	gray242		black243
	gray243		black244
	gray244		black245
	gray245		black246
	gray246		black247
	gray247		black248
	gray248		black249
	gray249		black24a
	gray24a		black24b
	gray24b		black24c
	gray24c		black24d
	gray24d		black24e
	gray24e		black24f
	gray24f		black250
	gray250		black251
	gray251		black252
	gray252		black253
	gray253		black254
	gray254		black255
	gray255		black256
	gray256		black257
	gray257		black258
	gray258		black259
	gray259		black25a
	gray25a		black25b
	gray25b		black25c
	gray25c		black25d
	gray25d		black25e
	gray25e		black25f
	gray25f		black260
	gray260		black261
	gray261		black262
	gray262		black263
	gray263		black264
	gray264		black265
	gray265		black266
	gray266		black267
	gray267		black268
	gray268		black269
	gray269		black26a
	gray26a		black26b
	gray26b		black26c
	gray26c		black26d
	gray26d		black26e
	gray26e		black26f
	gray26f		black270
	gray270		black271
	gray271		black272
	gray272		black273
	gray273		black274
	gray274		black275
	gray275		black276
	gray276		black277
	gray277		black278
	gray278		black279
	gray279		black27a
	gray27a		black27b
	gray27b		black27c
	gray27c		black27d
	gray27d		black27e
	gray27e		black27f
	gray27f		black280
	gray280		black281
	gray281		black282
	gray282		black283
	gray283		black284
	gray284		black285
	gray285		black286
	gray286		black287
	gray287		black288
	gray288		black289
	gray289		black28a
	gray28a		black28b
	gray28b		black28c
	gray28c		black28d
	gray28d		black28e
	gray28e		black28f
	gray28f		black290
	gray290		black291
	gray291		black292
	gray292		black293
	gray293		black294
	gray294		black295
	gray295		black296
	gray296		black297
	gray297		black298
	gray298		black299
	gray299		black29a
	gray29a		black29b
	gray29b		black29c
	gray29c</		

ggplot2 introduction

"ggplot2 is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base graphics and none of the bad parts. It takes care of many of the fiddly details that make plotting a hassle as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics."

-Hadley Wickham

ggplot2 introduction



- The "Tidyverse" is a collection of R packages designed using the Wickham philosophy
- **ggplot2** is the package in the Tidyverse collection used for plotting

```
# Install and load all at once
> install.packages("tidyverse")
> library("tidyverse")
```

```
# Install single package
> install.packages("ggplot2")
> library("ggplot2")
```

Wide vs. Long format

ggplot and other plotting packages like data in long format

```
> install.packages("reshape2")
> library("reshape2")
> my_data_long <- melt(my_data)
```

Wide-format

	Genotype	Tissue	Mapped	Unmapped
1	B73	shoot	0.82	0.06
2	Mo17	shoot	0.69	0.05
3	Mo17xB73	shoot	0.77	0.06
4	B73	root	0.85	0.05
5	Mo17	root	0.67	0.05
6	Mo17xB73	root	0.80	0.05
7	B73	internode	0.87	0.06
8	Mo17	internode	0.72	0.05
9	Mo17xB73	internode	0.80	0.05
10	B73	leaf	0.85	0.07
11	Mo17	leaf	0.69	0.06
12	Mo17xB73	leaf	0.75	0.06

Long-format

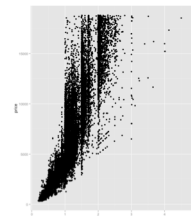
	Genotype	Tissue	variable	value
1	B73	shoot	Mapped	0.82
2	Mo17	shoot	Mapped	0.69
3	Mo17xB73	shoot	Mapped	0.77
4	B73	root	Mapped	0.85
5	Mo17	root	Mapped	0.67
6	Mo17xB73	root	Mapped	0.80
7	B73	internode	Mapped	0.87
8	Mo17	internode	Mapped	0.72
9	Mo17xB73	internode	Mapped	0.80
10	B73	leaf	Mapped	0.85
11	Mo17	leaf	Mapped	0.69
12	Mo17xB73	leaf	Mapped	0.75
13	B73	shoot	Unmapped	0.06
14	Mo17	shoot	Unmapped	0.05
15	Mo17xB73	shoot	Unmapped	0.06
16	B73	root	Unmapped	0.05
17	Mo17	root	Unmapped	0.05
18	Mo17xB73	root	Unmapped	0.05
19	B73	internode	Unmapped	0.06
20	Mo17	internode	Unmapped	0.05
21	Mo17xB73	internode	Unmapped	0.05
22	B73	leaf	Unmapped	0.07
23	Mo17	leaf	Unmapped	0.06
24	Mo17xB73	leaf	Unmapped	0.06

Quick ggplot

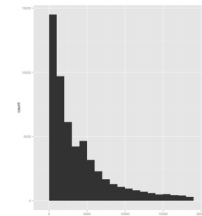
qplot is a quick plotting function within ggplot2 that is similar to the plot() function in base R

Best for getting a quick glance at your data

```
> qplot(diamonds$carat, diamonds$price)
```



```
> qplot(price, data = diamonds, geom="histogram", binwidth=1000)
```



ggplot philosophy

All ggplot2 plots start with `ggplot()`

- `data` – *data frame* containing data to be plotted
- `aes()` – *aesthetic* mappings to pass on the plot elements

```
> ggplot(mpg, aes(displ, hwy))
Error: No layers in plot
```

ggplot needs “layers” in the plot

- The geometric object (the type of plot you want) is a required layer

```
> ggplot(mpg, aes(displ, hwy)) + geom_point()
```

ggplot2 part of the tidyverse 3.1.0.9000

Layer: geoms

A layer combines data, aesthetic mapping, a geom (geometric object), a stat (statistical transformation), and a position adjustment. Typically, you will create layers using a `geom_` function, overriding the default position and stat if needed.

<code>geom_abline()</code>	<code>geom_hline()</code>	Reference lines: horizontal, vertical, and diagonal
<code>geom_vline()</code>		
<code>geom_bar()</code>	<code>geom_col()</code>	Bar charts
<code>stat_count()</code>		
<code>geom_bin2d()</code>	<code>stat_bin2d()</code>	Heatmap of 2d bin counts
<code>geom_blank()</code>		Draw nothing
<code>geom_boxplot()</code>	<code>stat_boxplot()</code>	A box and whiskers plot (in the style of Tukey)
<code>geom_contour()</code>	<code>stat_contour()</code>	2d contours of a 3d surface
<code>geom_count()</code>	<code>stat_sum()</code>	Count overlapping points
<code>geom_density()</code>	<code>stat_density()</code>	Smoothed density estimates
<code>geom_density_2d()</code>	<code>stat_density_2d()</code>	Contours of a 2d density estimate

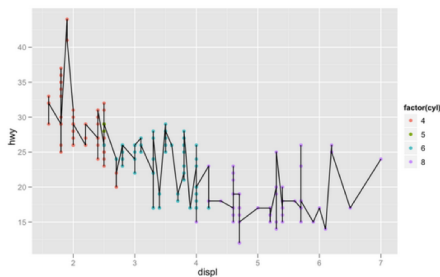
For a full <https://ggplot2.tidyverse.org/>

ggplot – intro to layers

Within a “geom” you can add new aesthetics

- `geom_line()` – add line to a point graph
- `aes(color = factor(cyl))` – color the points by cylinder type

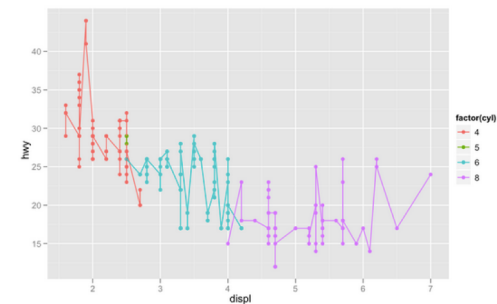
```
> ggplot(mpg, aes(displ, hwy)) + geom_point(aes(color = factor(cyl))) + geom_line()
```



ggplot – adding layers

We can also use the aesthetic mapping within the `ggplot` function to color both points and lines

```
> ggplot(mpg, aes(displ, hwy, color=factor(cyl))) + geom_point() + geom_line()
```



ggplot – additional customization

Many other “layers” can be added to customize a plot

Change the overall “look of plot”

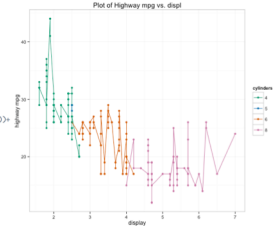
```
p <- ggplot(mpg, aes(displ, hwy, color=factor(cyl))) + geom_point() + geom_line() + theme_bw()
p + scale_colour_manual(name="cylinders", values=c("#009E73", "#0072B2", "#D55E00", "#CC79A7")) +
  labs(title="Plot of Highway mpg vs. displ", x="displacement", y="highway mpg", fill="my_legend")
```

Add custom line colors

Customize legend title

Add labels to the axis and title

```
#Plot intersection count per Mb of chromosome
ggplot(tempdata, aes(x=factor(chrom), y=pos1, z=value)) + stat_summary_2d(bins=30, fun="mean") +
  scale_x_discrete(name="Chromosome", labels=c("chr1", "chr2", "chr3", "chr4", "chr5", "chr6", "chr7", "chr8", "chr9", "chr10")) +
  labs(title = "Frequency Variants")
scale_y_continuous(name="Physical Position (Mb)", breaks=c(0, 50, 100, 150, 200, 250, 300)) +
  labs(title = "Frequency of Variants by Chromosome and Position for Trimmed Sequences", key = "Density")
theme(plot.title = element_text(size = 25, margin=margin(20, 20, 20, 20)),
      legend.key.size = unit(1, "cm"),
      legend.title = element_text(size = 20, margin=margin(20, 20, 20, 20)),
      axis.title.x = element_text(size = 20),
      axis.title.y = element_text(size = 20, angle = 45),
      axis.title.y = element_text(size = 20, margin=margin(20, 20, 20, 20)),
      axis.text.y = element_text(size = 20),
      panel.background = element_rect(fill = "white"))
ggsave("alignm_intersections.jpg", height=8, width=18)
```



Let's see some of this live...

Helpful ggplot Reference Page

<https://ggplot2.tidyverse.org/reference/>

2 matches **ggplot2** **geom_point** Done

part of the tidyverse
3.1.0.9000

Reference Articles News

Reference

Plot basics

All ggplot2 plots begin with a call to `ggplot()`, supplying default data and aesthetic mappings, specified by `aes()`. You then add layers, scales, coords and facets with `+`. To save a plot to disk, use `ggsave()`.

- `ggplot()` Create a new ggplot
- `aes()` Construct aesthetic mappings
- `+` (`ggplot`) `+` `geom` Add components to a plot
- `ggsave()` Save a ggplot (or other grid object) with sensible defaults
- `plot()` `quickplot()` Quick plot

Layer: geoms

A layer combines data, aesthetic mapping, a geom (geometric object), a stat (statistical transformation), and a position adjustment. Typically, you will create layers using a `geom_` function, overriding the

Contents

- Plot basics
- Layer: geoms
- Layer: stats
- Layer: position adjustment
- Layer: annotations
- Aesthetics
- Scales
- Guides: axes and legends
- Facetting
- Facetting: labels
- Coordinate systems
- Themes
- Programming with ggplot2
- Extending ggplot2
- Vector helpers
- Data
- Autoplot and fortify

R Markdown in R Studio

- Tutorial at <http://rmarkdown.rstudio.com/lesson-1.html>

Hirsch/Hirsch/Springer RNAseq Analysis Pipeline

Cory Hirsch
September 6, 2016

These notes are for general RNAseq analysis methods used in the Hirsch, Hirsch, and Springer Lab groups. It isn't the only way to do things and each project might require tweaking the methods.

Outline of steps

- Download reads from SRA to HIS
- Extract fastq format from .sra format files
- Check the quality of raw sequence files (FastQC)
- Remove adapter sequence and low quality bases from 3' end (Cutadapt)
- Check quality of 'cleaned' reads (FastQC)
- Index the reference (BWA)
- Align Illumina reads to the genome (BWA)
- Determine the alignment metrics (samtools)
- Determine the number of reads aligning to each feature (HTSeq)
- Make an expression matrix consisting of all the samples into 1 file (Perl Script)
- Turn read counts into rpm values (Perl Script)
- Determine differentially expressed genes (DESeq2)

Shell Script Example

Example of standard portion of shell script for submitting jobs to HIS

```
#!/bin/bash
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
```

Example of portion of shell script to call information from another file

```
#!/bin/bash
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
# Set the path to the directory where the script is located
HIS_PATH="/home/hirsch/seqdata/his"
```

- Easy to keep organized
- Images contained within notes (if generated in R!)
- Flat file that is rendered into a .html file to view