Quick Wins in R

— The Ultimate Beginner's Cheat Sheet —

Getting started

1. Download RStudio from RStudio.com, and use it to write and run your R scripts
2. Start a new project and load in data files for easy access
3. To install ggplot2: go to the *Packages* tab → Click *Install* → Type *ggplot2* → Click *Install*

Data at first sight

Import data

```
# Read data from a file into a data.frame (Use sep="," for CSV files)
my_data <- read.csv("pokemon.tsv", sep="\t", header=TRUE)
```

Inspect the data

```
# Preview the top of the file
head(my_data)
```

```
# View the column names
names(my_data)
```

```
# Count number of rows and columns
dim(my_data)
```

Basic statistics

```
# Basic statistics
summary(my_data)
```

```
# Stats for a categorical column:
names(my_data$Type_1)
```

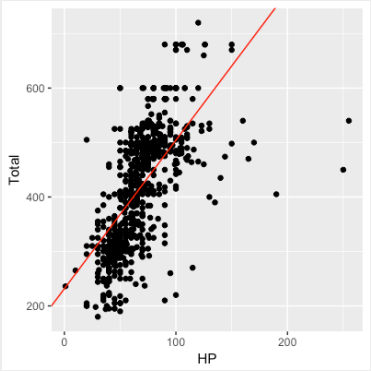
```
# Stats for a numerical column
summary(my_data$HP)
mean(my_data$HP); sd(my_data$HP)
```

Correlation

```
# Calculate correlation between HP and Total
r <- cor(my_data$HP, my_data$Total)
r^2
```

Linear regression

```
# Create a fit where y=Total, x=HP:
fit <- lm(Total ~ HP, data=my_data)

# y = mx + b
b <- fit$coefficients[1]
m <- fit$coefficients[2]

# Draw a scatterplot with a fit line
ggplot(my_data, aes(x=HP, y=Total)) +
  geom_point() +
  geom_abline(slope=m, intercept=b, color="red")
```



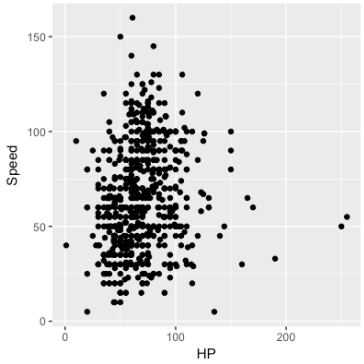
This cheat sheet comes with a free course! Get it at OMGenomics.com/R

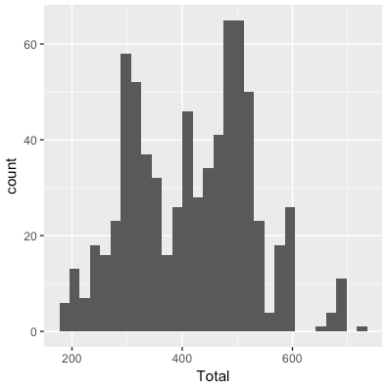
First: Load the ggplot2 library

```
library(ggplot2)
```

Scatter plot

```
# Simple scatter plot
ggplot(my_data, aes(x=HP, y=Speed)) + geom_point()

# Color points by another column
ggplot(my_data, aes(x=HP, y=Speed, color=Type_1)) + geom_point()
ggplot(my_data, aes(x=HP, y=Speed, color=Attack)) + geom_point()
```



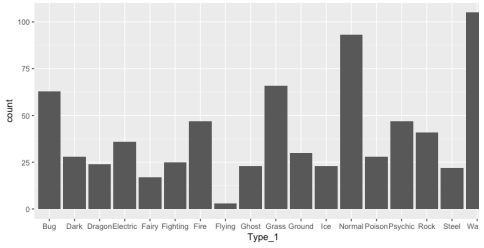


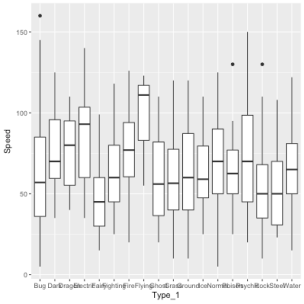
Histogram

```
# Set the total number of bins
ggplot(my_data, aes(x=HP)) + geom_histogram(bins=20)

# Set the binwidth: good for discrete numbers
ggplot(my_data, aes(x=HP)) + geom_histogram(binwidth=1)
```

Bar chart

```
# Simple bar chart
ggplot(my_data, aes(x=Type_1)) + geom_bar()

# Color bars by another column
ggplot(my_data, aes(x=Type_1, fill=isLegendary)) + geom_bar()
```





Box plot

```
# Simple box plot
ggplot(my_data, aes(x=Type_1, y=Speed)) + geom_boxplot()

# Splitting and coloring by another column
ggplot(my_data, aes(x=Type_1, y=Speed, fill=hasGender)) + geom_boxplot()
```

Save an image of the plot as a file

```
png("my_plot.png", width=1000, height=1000, res=100)
  # ggplot code here
dev.off()
# Or replace png with: tiff, jpeg, pdf
```