Class 5: Data Viz with ggplot

Michael A18135877

Today we ar eexploring the **gggplot** package and how to make nice figures in R.

There are lots of ways to make figures and plot in R. These include:

- so called "base" R
- and add on packages like **ggplot2**

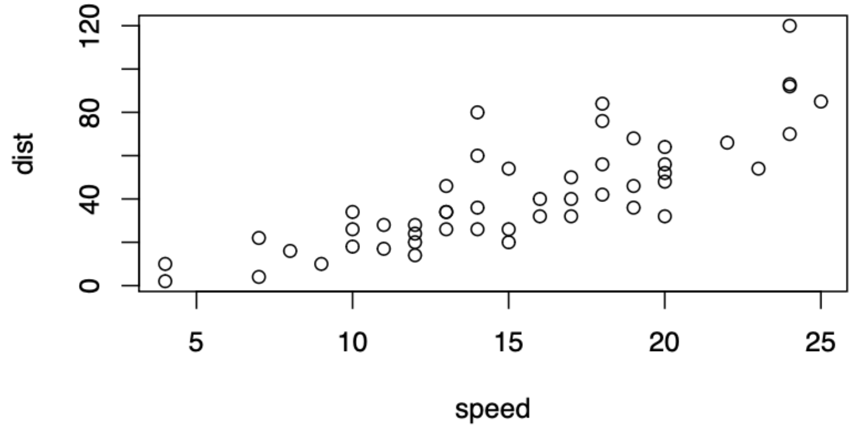
Here is a simple "base" R plot.

```r
head(cars)
```

```
  speed dist
1     4    2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
```

We can simply pass this to the `plot()` function.

```r
plot(cars)
```



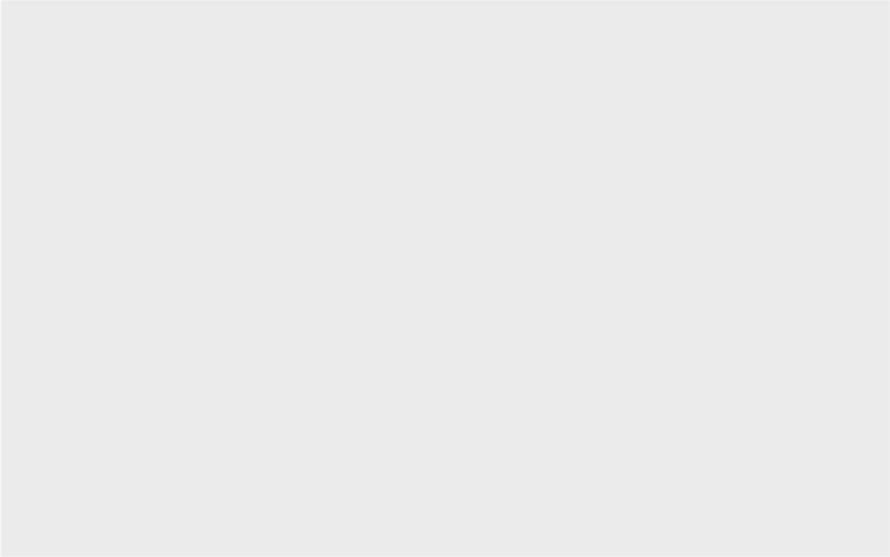
> Key-point: Base R is quick but not so nice looking in some folks eyes.

Let's see how we can plot this with **ggplot2**...

1st I need to install this add-on package. For this we use the `install.packages()` function - **WE DO THIS IN THE CONSOLE, NOT our report**. This is a one time only deal.

2nd We need to load the package with `library()` function every time we want to use it.

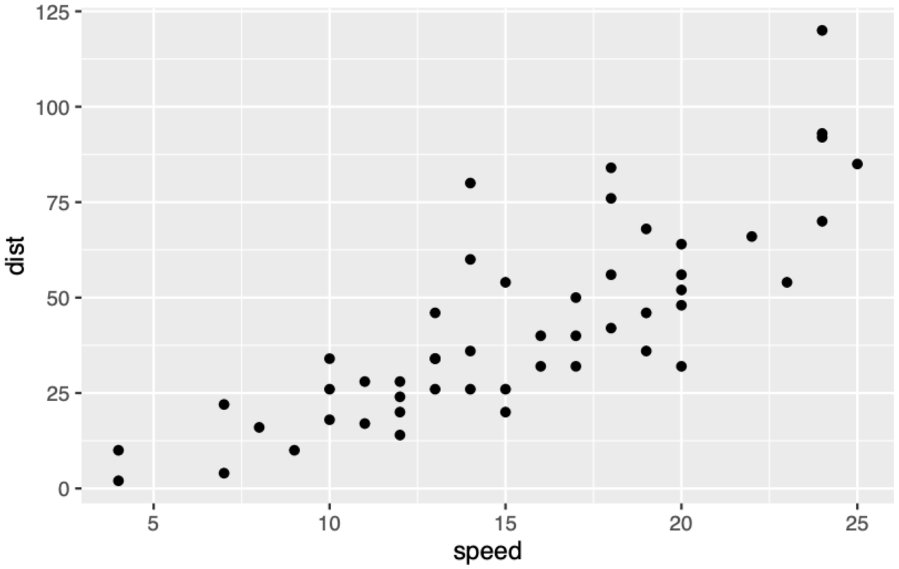
```
library(ggplot2)
ggplot(cars)
```



Every ggplot is composed of at least 3 layers:

- **data** (i.e a data.frame with the things you want to plot),
- aesthitics **aes()** that map the columns of data to your plot features (i.e. aesthitics)
- geoms like **geom_point()** that sort how the plot appears

```
ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point()
```



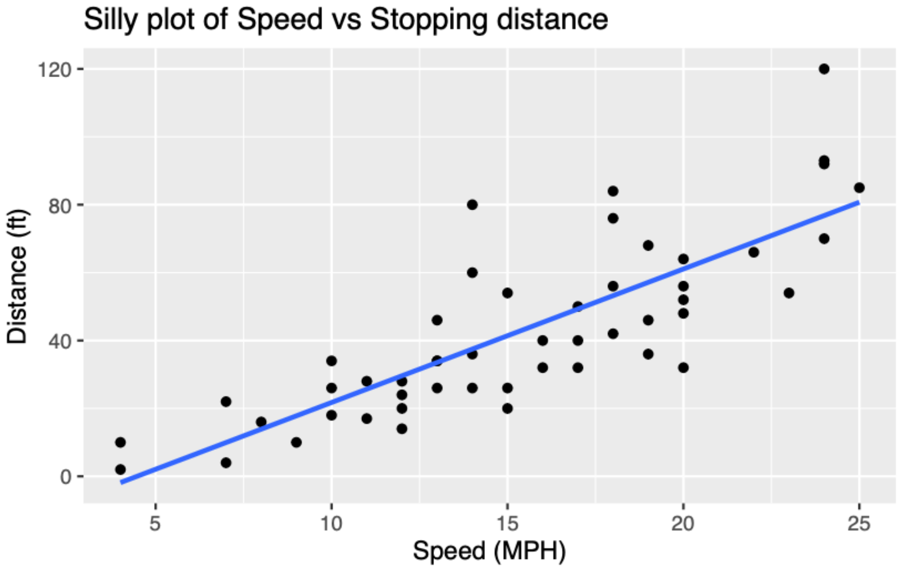
> Key point: For simple “canned” graphs base R is quicker but as things get more custom and elaborate then ggplot wins out…

Let’s add more layers to our ggplot

Add a line showing the relationship between x and y Add a title Add custom axis labels “Speed (MPH)” and “Distance (ft)” Chnage the theme…

```r
ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point() +
  geom_smooth(method="lm", se=F) +
  labs(title="Silly plot of Speed vs Stopping distance") +
  labs(x="Speed (MPH)", y="Distance (ft)")
```

```
`geom_smooth()` using formula = 'y ~ x'
```



```
theme_bw()
```

```
<theme> List of 144
 $ line                              : <ggplot2::element_line>
  ..@ colour       : chr "black"
  ..@ linewidth    : num 0.5
  ..@ linetype     : num 1
  ..@ lineend      : chr "butt"
  ..@ linejoin     : chr "round"
  ..@ arrow        : logi FALSE
  ..@ arrow.fill   : chr "black"
  ..@ inherit.blank: logi TRUE
 $ rect                              : <ggplot2::element_rect>
  ..@ fill         : chr "white"
  ..@ colour       : chr "black"
  ..@ linewidth    : num 0.5
  ..@ linetype     : num 1
  ..@ linejoin     : chr "round"
  ..@ inherit.blank: logi TRUE
 $ text                              : <ggplot2::element_text>
  ..@ family       : chr ""
  ..@ face         : chr "plain"
  ..@ italic       : chr NA
  ..@ fontweight   : num NA
  ..@ fontwidth    : num NA
  ..@ colour       : chr "black"
  ..@ size         : num 11
  ..@ hjust        : num 0.5
  ..@ vjust        : num 0.5
  ..@ angle        : num 0
  ..@ lineheight   : num 0.9
  ..@ margin       : <ggplot2::margin> num [1:4] 0 0 0 0
  ..@ debug        : logi FALSE
  ..@ inherit.blank: logi TRUE
 $ title                            : <ggplot2::element_text>
  ..@ family       : NULL
  ..@ face         : NULL
  ..@ italic       : chr NA
  ..@ fontweight   : num NA
  ..@ fontwidth    : num NA
  ..@ colour       : NULL
  ..@ size         : NULL
  ..@ hjust        : NULL
  ..@ vjust        : NULL
  ..@ angle        : NULL
  ..@ lineheight   : NULL
  ..@ margin       : NULL
  ..@ debug        : NULL
  ..@ inherit.blank: logi TRUE
 $ point                            : <ggplot2::element_point>
  ..@ colour       : chr "black"
  ..@ shape        : num 19
  ..@ size         : num 1.5
  ..@ fill         : chr "white"
  ..@ stroke       : num 0.5
  ..@ inherit.blank: logi TRUE
 $ polygon                          : <ggplot2::element_polygon>
  ..@ fill         : chr "white"
  ..@ colour       : chr "black"
  ..@ linewidth    : num 0.5
  ..@ linetype     : num 1
  ..@ linejoin     : chr "round"
  ..@ inherit.blank: logi TRUE
 $ geom                             : <ggplot2::element_geom>
  ..@ ink        : chr "black"
 ..@ paper      : chr "white"
 ..@ accent     : chr "#3366FF"
 ..@ linewidth  : num 0.5
 ..@ borderwidth: num 0.5
 ..@ linetype   : int 1
 ..@ bordertype : int 1
 ..@ family     : chr ""
 ..@ fontsize   : num 3.87
 ..@ pointsize  : num 1.5
 ..@ pointshape : num 19
 ..@ colour     : NULL
 ..@ fill       : NULL
$ spacing                          : 'simpleUnit' num 5.5points
 ..- attr(*, "unit")= int 8
$ margins                          : <ggplot2::margin> num [1:4] 5.5 5.5 5.5 5.5
$ aspect.ratio                     : NULL
$ axis.title                       : NULL
$ axis.title.x                     : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : NULL
 ..@ hjust        : NULL
 ..@ vjust        : num 1
 ..@ angle        : NULL
 ..@ lineheight   : NULL
 ..@ margin       : <ggplot2::margin> num [1:4] 2.75 0 0 0
 ..@ debug        : NULL
 ..@ inherit.blank: logi TRUE
$ axis.title.x.top                 : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : NULL
 ..@ hjust        : NULL
 ..@ vjust        : num 0
 ..@ angle        : NULL
```

```
  ..@ lineheight   : NULL
  ..@ margin       : <ggplot2::margin> num [1:4] 0 0 2.75 0
  ..@ debug        : NULL
  ..@ inherit.blank: logi TRUE
 $ axis.title.x.bottom            : NULL
 $ axis.title.y                   : <ggplot2::element_text>
  ..@ family       : NULL
  ..@ face         : NULL
  ..@ italic       : chr NA
  ..@ fontweight   : num NA
  ..@ fontwidth    : num NA
  ..@ colour       : NULL
  ..@ size         : NULL
  ..@ hjust        : NULL
  ..@ vjust        : num 1
  ..@ angle        : num 90
  ..@ lineheight   : NULL
  ..@ margin       : <ggplot2::margin> num [1:4] 0 2.75 0 0
  ..@ debug        : NULL
  ..@ inherit.blank: logi TRUE
 $ axis.title.y.left              : NULL
 $ axis.title.y.right             : <ggplot2::element_text>
  ..@ family       : NULL
  ..@ face         : NULL
  ..@ italic       : chr NA
  ..@ fontweight   : num NA
  ..@ fontwidth    : num NA
  ..@ colour       : NULL
  ..@ size         : NULL
  ..@ hjust        : NULL
  ..@ vjust        : num 1
  ..@ angle        : num -90
  ..@ lineheight   : NULL
  ..@ margin       : <ggplot2::margin> num [1:4] 0 0 0 2.75
  ..@ debug        : NULL
  ..@ inherit.blank: logi TRUE
 $ axis.text                      : <ggplot2::element_text>
  ..@ family       : NULL
  ..@ face         : NULL
  ..@ italic       : chr NA
  ..@ fontweight   : num NA
  ..@ fontwidth    : num NA
  ..@ colour       : chr "#4D4D4DFF"
 ..@ size         : 'rel' num 0.8
 ..@ hjust        : NULL
 ..@ vjust        : NULL
 ..@ angle        : NULL
 ..@ lineheight   : NULL
 ..@ margin       : NULL
 ..@ debug        : NULL
 ..@ inherit.blank: logi TRUE
$ axis.text.x                     : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : NULL
 ..@ hjust        : NULL
 ..@ vjust        : num 1
 ..@ angle        : NULL
 ..@ lineheight   : NULL
 ..@ margin       : <ggplot2::margin> num [1:4] 2.2 0 0 0
 ..@ debug        : NULL
 ..@ inherit.blank: logi TRUE
$ axis.text.x.top                 : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : NULL
 ..@ hjust        : NULL
 ..@ vjust        : num 0
 ..@ angle        : NULL
 ..@ lineheight   : NULL
 ..@ margin       : <ggplot2::margin> num [1:4] 0 0 2.2 0
 ..@ debug        : NULL
 ..@ inherit.blank: logi TRUE
$ axis.text.x.bottom              : NULL
$ axis.text.y                     : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : NULL
 ..@ hjust        : num 1
 ..@ vjust        : NULL
 ..@ angle        : NULL
 ..@ lineheight   : NULL
 ..@ margin       : <ggplot2::margin> num [1:4] 0 2.2 0 0
 ..@ debug        : NULL
 ..@ inherit.blank: logi TRUE
$ axis.text.y.left                : NULL
$ axis.text.y.right               : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : NULL
 ..@ hjust        : num 0
 ..@ vjust        : NULL
 ..@ angle        : NULL
 ..@ lineheight   : NULL
 ..@ margin       : <ggplot2::margin> num [1:4] 0 0 0 2.2
 ..@ debug        : NULL
 ..@ inherit.blank: logi TRUE
$ axis.text.theta                 : NULL
$ axis.text.r                     : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : NULL
 ..@ hjust        : num 0.5
 ..@ vjust        : NULL
 ..@ angle        : NULL
 ..@ lineheight   : NULL
 ..@ margin       : <ggplot2::margin> num [1:4] 0 2.2 0 2.2
 ..@ debug        : NULL
 ..@ inherit.blank: logi TRUE
```

```
$ axis.ticks                      : <ggplot2::element_line>
 ..@ colour       : chr "#333333FF"
 ..@ linewidth    : NULL
 ..@ linetype     : NULL
 ..@ lineend      : NULL
 ..@ linejoin     : NULL
 ..@ arrow        : logi FALSE
 ..@ arrow.fill   : chr "#333333FF"
 ..@ inherit.blank: logi TRUE
$ axis.ticks.x                    : NULL
$ axis.ticks.x.top                : NULL
$ axis.ticks.x.bottom             : NULL
$ axis.ticks.y                    : NULL
$ axis.ticks.y.left               : NULL
$ axis.ticks.y.right              : NULL
$ axis.ticks.theta                : NULL
$ axis.ticks.r                    : NULL
$ axis.minor.ticks.x.top          : NULL
$ axis.minor.ticks.x.bottom       : NULL
$ axis.minor.ticks.y.left         : NULL
$ axis.minor.ticks.y.right        : NULL
$ axis.minor.ticks.theta          : NULL
$ axis.minor.ticks.r              : NULL
$ axis.ticks.length               : 'rel' num 0.5
$ axis.ticks.length.x             : NULL
$ axis.ticks.length.x.top         : NULL
$ axis.ticks.length.x.bottom      : NULL
$ axis.ticks.length.y             : NULL
$ axis.ticks.length.y.left        : NULL
$ axis.ticks.length.y.right       : NULL
$ axis.ticks.length.theta         : NULL
$ axis.ticks.length.r             : NULL
$ axis.minor.ticks.length         : 'rel' num 0.75
$ axis.minor.ticks.length.x       : NULL
$ axis.minor.ticks.length.x.top   : NULL
$ axis.minor.ticks.length.x.bottom: NULL
$ axis.minor.ticks.length.y       : NULL
$ axis.minor.ticks.length.y.left  : NULL
$ axis.minor.ticks.length.y.right : NULL
$ axis.minor.ticks.length.theta   : NULL
$ axis.minor.ticks.length.r       : NULL
$ axis.line                       : <ggplot2::element_blank>
$ axis.line.x                     : NULL
$ axis.line.x.top                 : NULL
$ axis.line.x.bottom              : NULL
$ axis.line.y                     : NULL
$ axis.line.y.left                : NULL
$ axis.line.y.right               : NULL
$ axis.line.theta                 : NULL
$ axis.line.r                     : NULL
$ legend.background               : <ggplot2::element_rect>
 ..@ fill         : NULL
 ..@ colour       : logi NA
 ..@ linewidth    : NULL
 ..@ linetype     : NULL
 ..@ linejoin     : NULL
 ..@ inherit.blank: logi TRUE
$ legend.margin                   : NULL
$ legend.spacing                  : 'rel' num 2
$ legend.spacing.x                : NULL
$ legend.spacing.y                : NULL
$ legend.key                      : NULL
$ legend.key.size                 : 'simpleUnit' num 1.2lines
 ..- attr(*, "unit")= int 3
$ legend.key.height               : NULL
$ legend.key.width                : NULL
$ legend.key.spacing              : NULL
$ legend.key.spacing.x            : NULL
$ legend.key.spacing.y            : NULL
$ legend.key.justification        : NULL
$ legend.frame                    : NULL
$ legend.ticks                    : NULL
$ legend.ticks.length             : 'rel' num 0.2
$ legend.axis.line                : NULL
$ legend.text                     : <ggplot2::element_text>
 ..@ family       : NULL
 ..@ face         : NULL
 ..@ italic       : chr NA
 ..@ fontweight   : num NA
 ..@ fontwidth    : num NA
 ..@ colour       : NULL
 ..@ size         : 'rel' num 0.8
 ..@ hjust        : NULL
 ..@ vjust        : NULL
 ..@ angle        : NULL
 ..@ lineheight   : NULL
```

```
  ..@ margin       : NULL
  ..@ debug        : NULL
  ..@ inherit.blank: logi TRUE
 $ legend.text.position            : NULL
 $ legend.title                    : <ggplot2::element_text>
  ..@ family       : NULL
  ..@ face         : NULL
  ..@ italic       : chr NA
  ..@ fontweight   : num NA
  ..@ fontwidth    : num NA
  ..@ colour       : NULL
  ..@ size         : NULL
  ..@ hjust        : num 0
  ..@ vjust        : NULL
  ..@ angle        : NULL
  ..@ lineheight   : NULL
  ..@ margin       : NULL
  ..@ debug        : NULL
  ..@ inherit.blank: logi TRUE
 $ legend.title.position           : NULL
 $ legend.position                 : chr "right"
 $ legend.position.inside          : NULL
 $ legend.direction                : NULL
 $ legend.byrow                    : NULL
 $ legend.justification            : chr "center"
 $ legend.justification.top        : NULL
 $ legend.justification.bottom     : NULL
 $ legend.justification.left       : NULL
 $ legend.justification.right      : NULL
 $ legend.justification.inside     : NULL
  [list output truncated]
 @ complete: logi TRUE
 @ validate: logi TRUE
```

Going further

Read some expression data

```r
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
head(genes)
```

```
        Gene Condition1 Condition2      State
1      A4GNT -3.6808610 -3.4401355 unchanging
2       AAAS  4.5479580  4.3864126 unchanging
3      AASDH  3.7190695  3.4787276 unchanging
4       AATF  5.0784720  5.0151916 unchanging
5       AATK  0.4711421  0.5598642 unchanging
6 AB015752.4 -3.6808610 -3.5921390 unchanging
```

> Q1. How many genes are in this wee dataset?

```r
nrow(genes)
```

```
[1] 5196
```

> Q2. How many "up" regulated genes are there?

```r
table(genes$State)
```

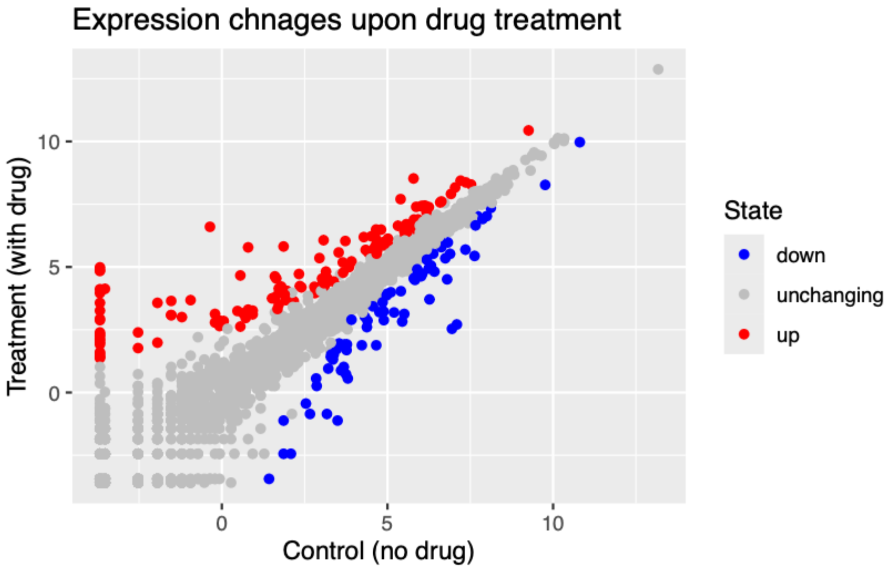
```

      down unchanging         up 
        72       4997        127 
```

A useful function for counting up occurances of things in a vector is the `table()` function.

Make a v1 figure

```r
p<-ggplot(genes) +
  aes(x= Condition1,
      y= Condition2,
      col=State) +
  geom_point() +
  labs(title= "Expression chnages upon drug treatment",
       x="Control (no drug)",
       y="Treatment (with drug)")
  p + scale_colour_manual(values = c("blue", "gray","red"))
```



##More PLotting library(gapminder)

```
# File location online
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.ts

gapminder <- read.delim(url)
```

Lets have a wee peak

```
head(gapminder, 3)
```

```
      country continent year lifeExp      pop gdpPercap
1 Afghanistan      Asia 1952  28.801  8425333  779.4453
2 Afghanistan      Asia 1957  30.332  9240934  820.8530
3 Afghanistan      Asia 1962  31.997 10267083  853.1007
```

> Q4. How many different country values are in this dataset?

```
nrow(gapminder)
```

```
[1] 1704
```

```
length(table(gapminder$country))
```

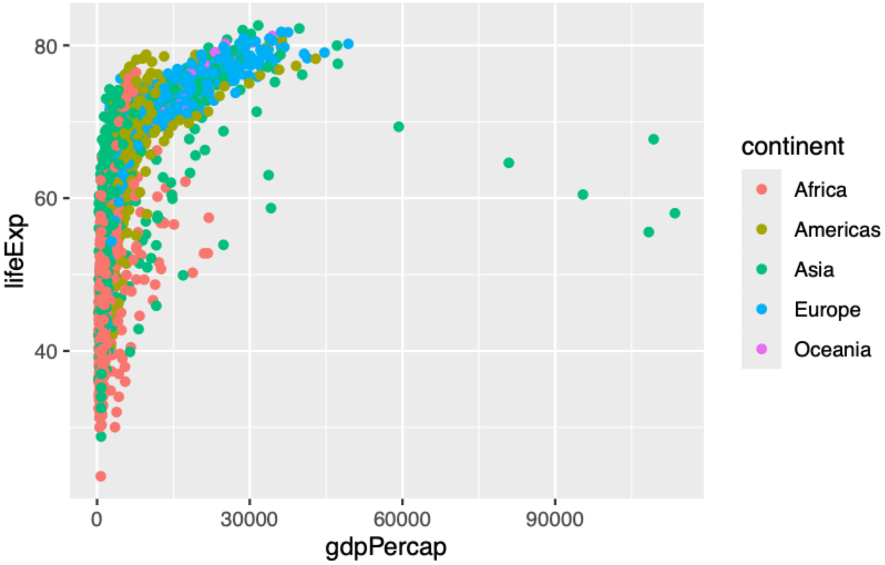
```
[1] 142
```

> Q5. How many different continent values are in this dataset.

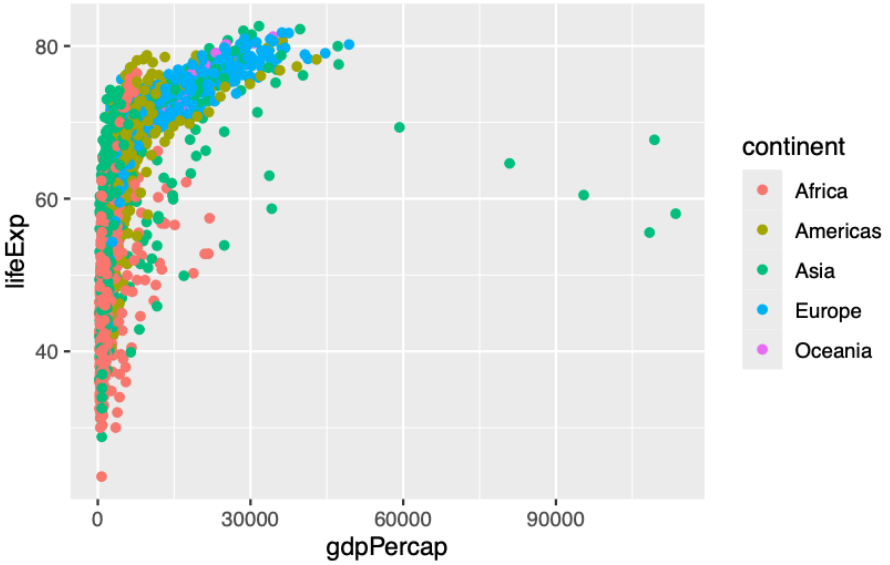
```
unique(gapminder$continent)
```

```
[1] "Asia"     "Europe"   "Africa"   "Americas" "Oceania"
```

```
ggplot(gapminder) +
  aes(gdpPercap, lifeExp, color=continent) +
  geom_point()
```



```
ggplot(gapminder) +
  aes(gdpPercap, lifeExp, color=continent, label=country) +
  geom_point()
```

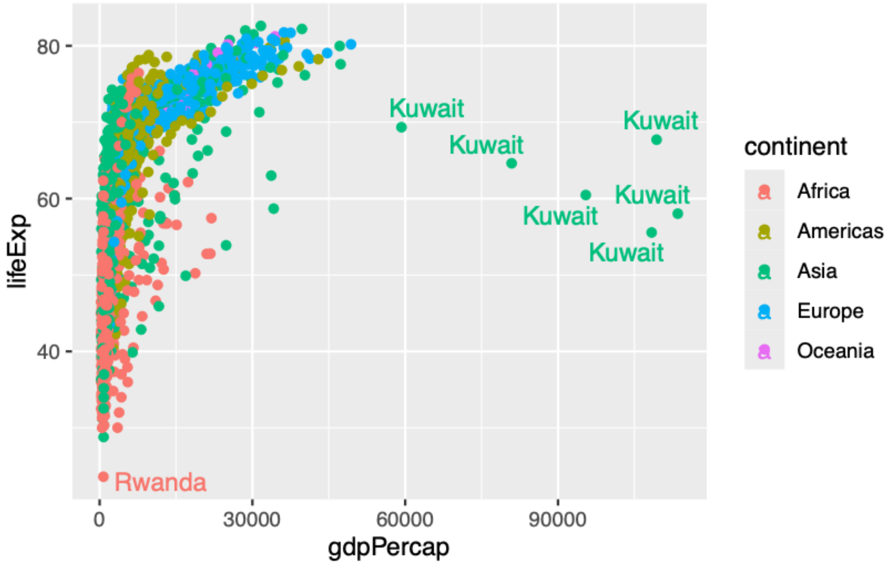


I can use the **ggrepel** package to make more sensible labels here.

```
library(ggrepel)

ggplot(gapminder) +
  aes(gdpPercap, lifeExp, color=continent, label=country) +
  geom_point() +
  geom_text_repel()
```

```
Warning: ggrepel: 1697 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



I want a separate panel per continent

```
ggplot(gapminder) +
  aes(gdpPercap, lifeExp, color=continent, label=country) +
  geom_point() +
  geom_text_repel() +
  facet_wrap(~continent)
```

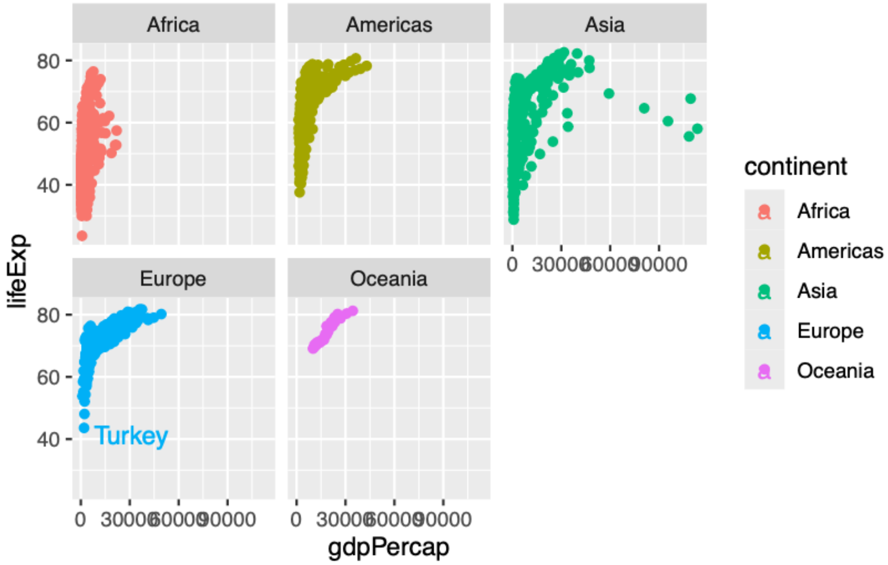
```
Warning: ggrepel: 624 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```

```
Warning: ggrepel: 359 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```

```
Warning: ggrepel: 300 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```

```
Warning: ggrepel: 24 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```

```
Warning: ggrepel: 396 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



ggplot2 offers several advantages over base R plots:

1. **Layered Grammar**: ggplot2 uses a consistent, layered approach for building plots, making it easier to add or modify elements (data, aesthetics, geoms) step by step [1], [2], [3], [5], [4].
2. **Publication Quality**: ggplot2 produces visually appealing, publication-quality figures with sensible defaults, which are harder to achieve with base R without extensive tweaking [1], [2], [3], [4].
3. **Customization**: Customizing complex plots is more straightforward in ggplot2. Adding legends, themes, or combining multiple plots is easier and less error-prone than in base R [1], [2], [3].
4. **Consistency**: The syntax and logic are consistent across different plot types, reducing the need to learn many separate functions as in base R [1], [2], [3], [5].
5. **Extensibility**: ggplot2 is part of a larger ecosystem, allowing integration with other packages for advanced graphics and analysis [1].

What do you think is the most important advantage for your own work?