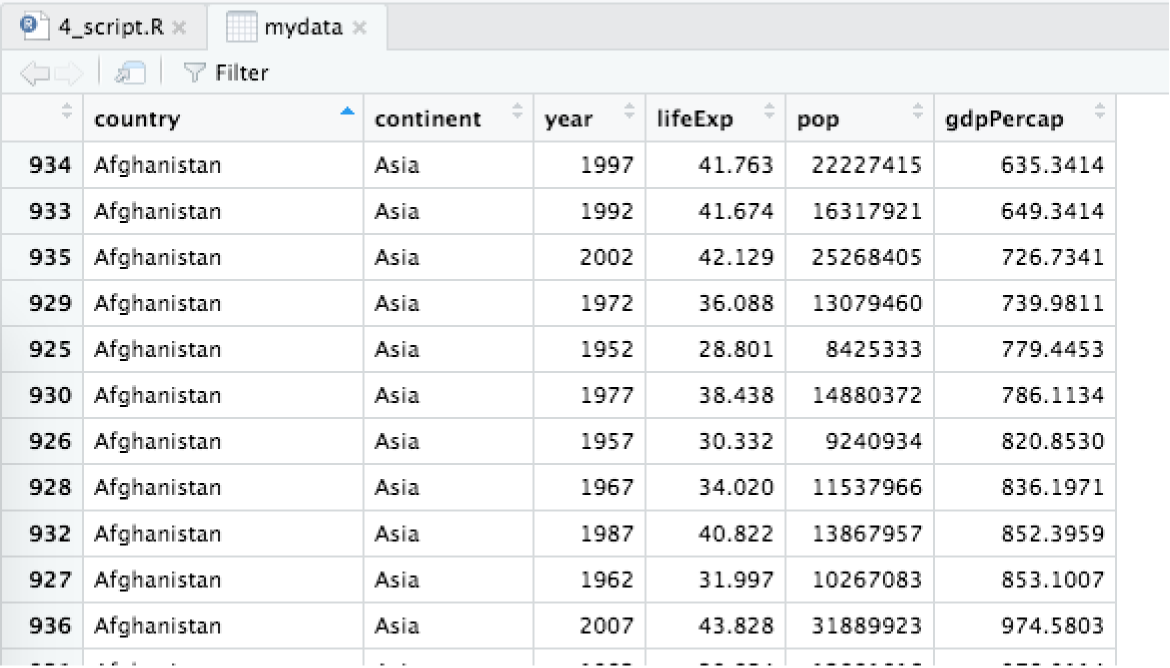
An introduction by Riinu Pius

Artwork by @allison_horst

We will be using the gapminder dataset (https://www.gapminder.org/):



	country	continent	year	lifeExp	pop	gdpPercap
934	Afghanistan	Asia	1997	41.763	22227415	635.3414
933	Afghanistan	Asia	1992	41.674	16317921	649.3414
935	Afghanistan	Asia	2002	42.129	25268405	726.7341
929	Afghanistan	Asia	1972	36.088	13079460	739.9811
925	Afghanistan	Asia	1952	28.801	8425333	779.4453
930	Afghanistan	Asia	1977	38.438	14880372	786.1134
926	Afghanistan	Asia	1957	30.332	9240934	820.8530
928	Afghanistan	Asia	1967	34.020	11537966	836.1971
932	Afghanistan	Asia	1987	40.822	13867957	852.3959
927	Afghanistan	Asia	1962	31.997	10267083	853.1007
936	Afghanistan	Asia	2007	43.828	31889923	974.5803

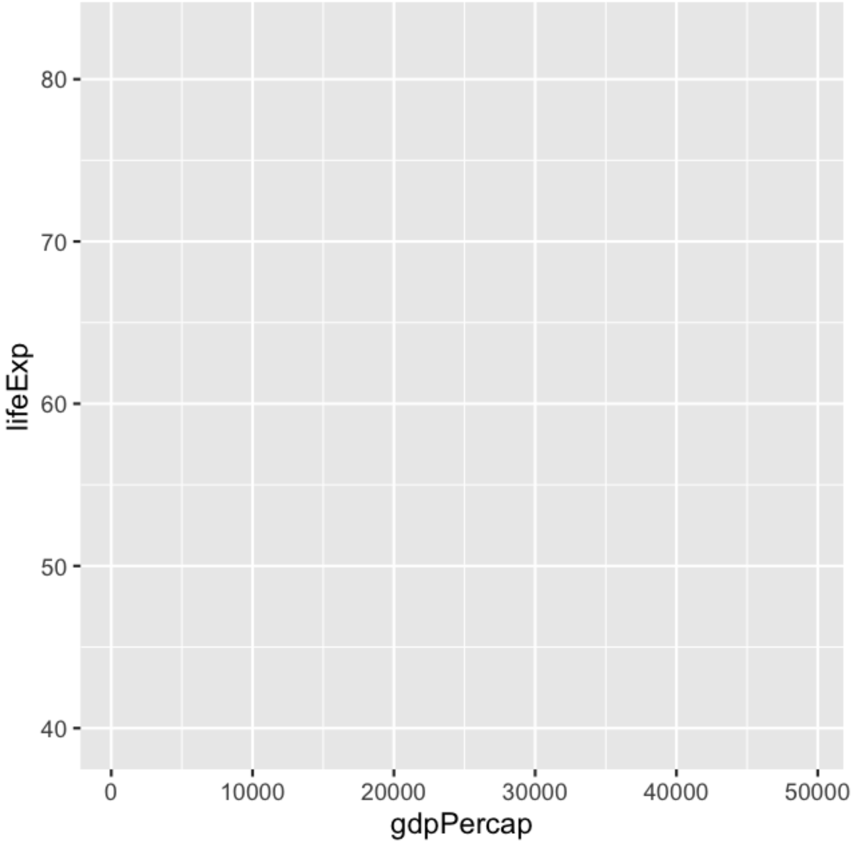
```
gapdata$year %>%   unique()
```

```
##  [1] 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997
## [11] 2002 2007
```

`ggplot()` initialises a canvas

Important: Before `ggplot()` use the pipe (`%>%`); after `ggplot()` use plus (`+`):

```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y=lifeExp))
```

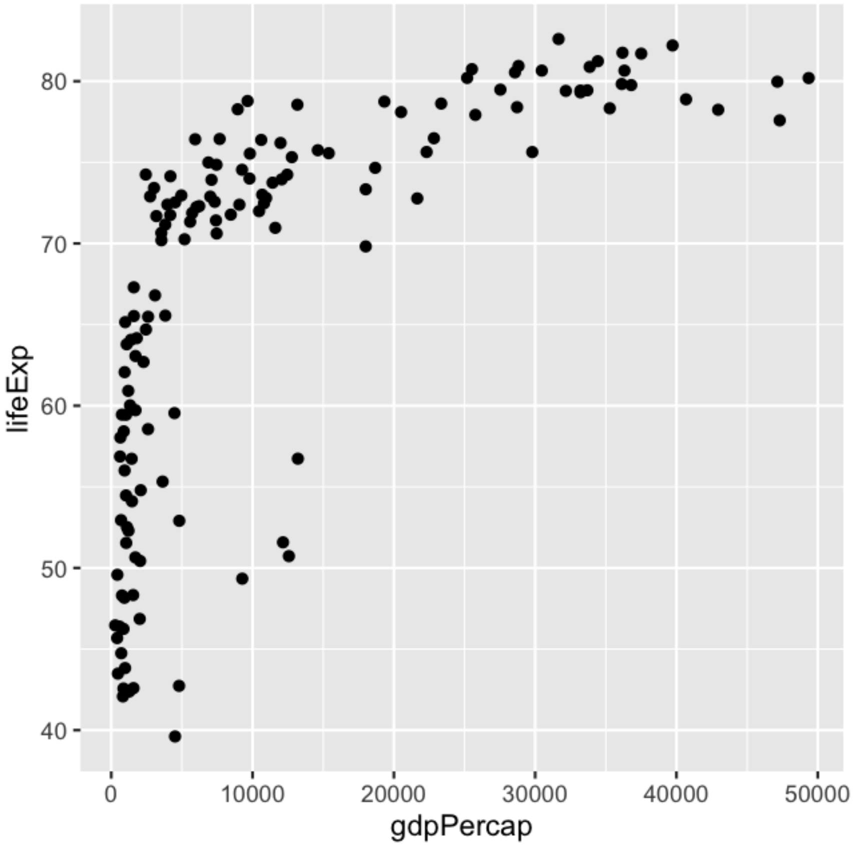


`filter()` is often useful before `ggplot()`

...a canvas to which we can add geoms

geom stands for geometrical. Here we've added `geom_point()`

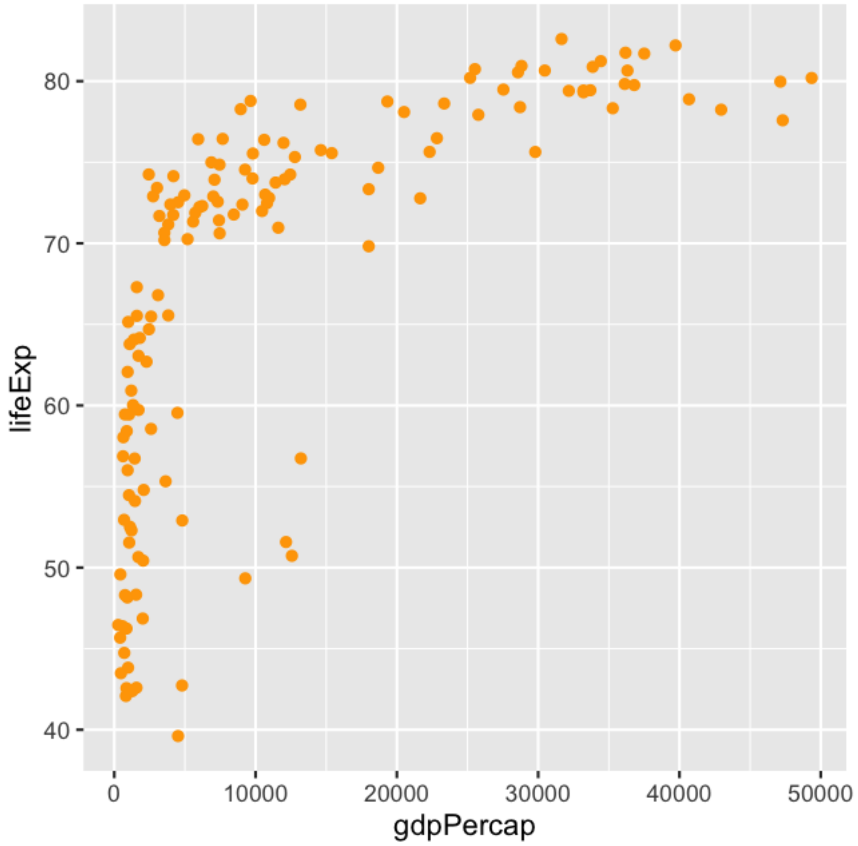
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp)) +
  geom_point()
```



specifying visual properties outside `aes()`

Let's change the colour of the points: `colour = "orange"`

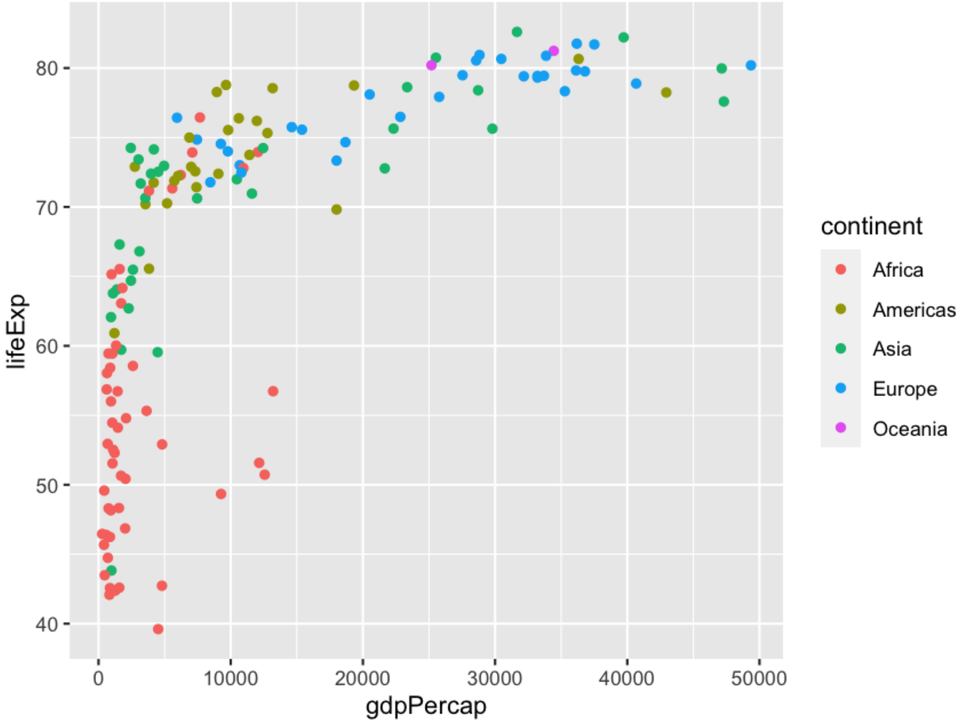
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp)) +
  geom_point(colour = "orange")
```



specifying visual properties inside `aes()`

Let's use the variable `continent` to colour the points by (`colour = continent` inside `aes()`):

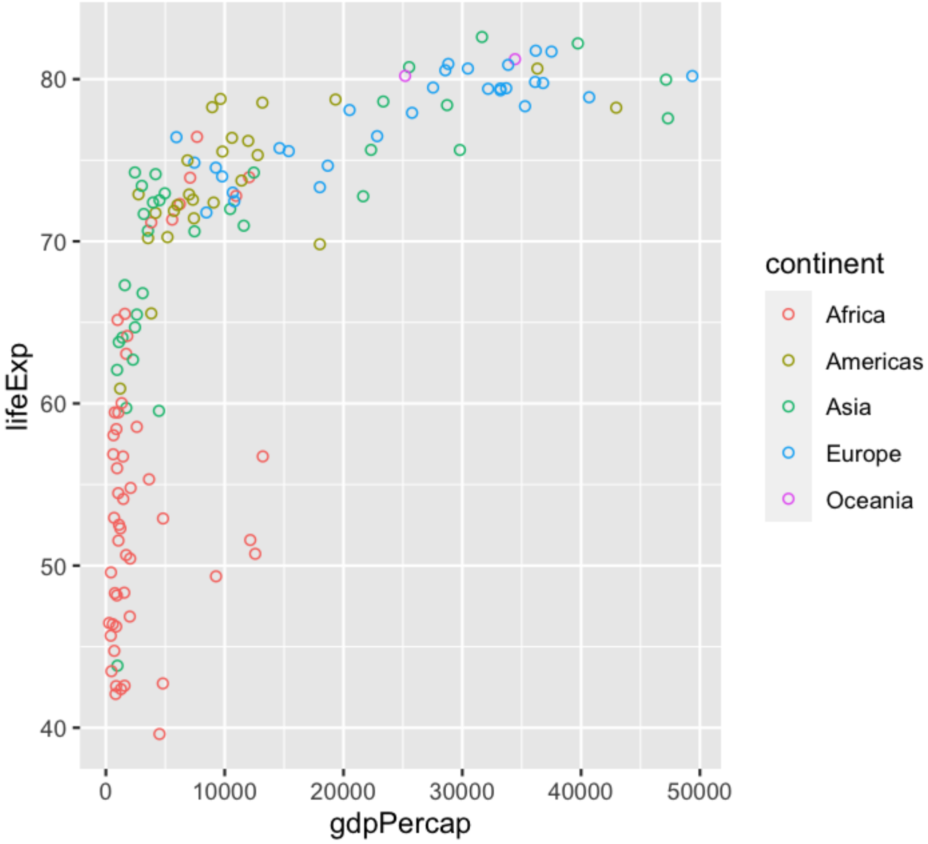
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
                        colour = continent)) +
  geom_point()
```



There is no limit* to the `aes()` you can include

In addition to colouring the points by continent, we can size them by population:

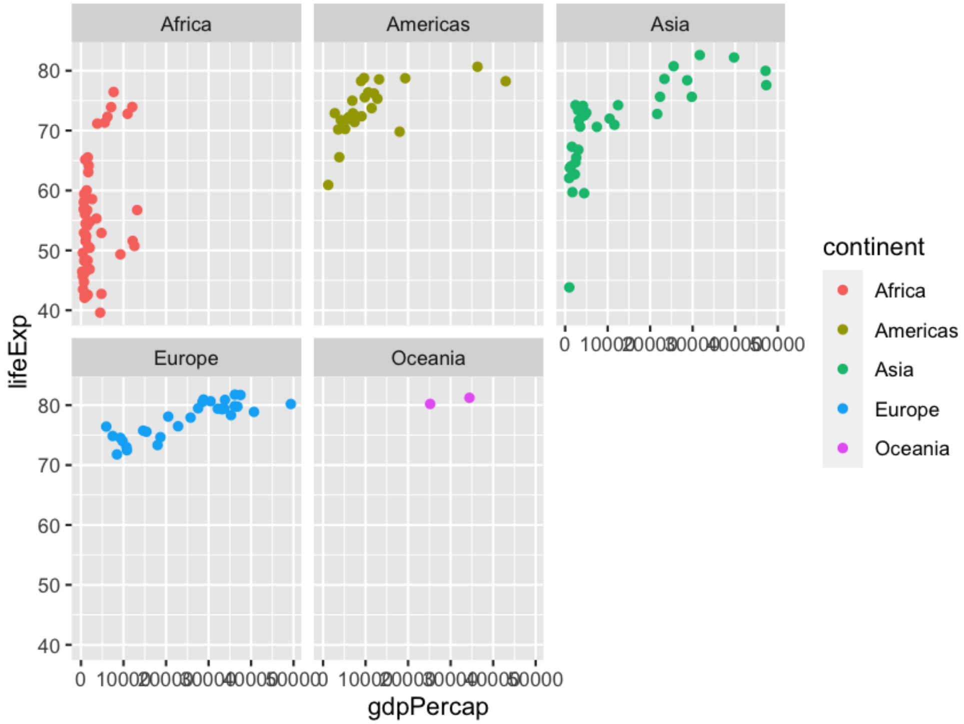
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
                      colour = continent)) +
  geom_point(shape = 1)
```



* there is a limit. Press F1 on `geom_point()` to see the list of aesthetics

From 1 plot to 5 with `facet_wrap(~continent)`

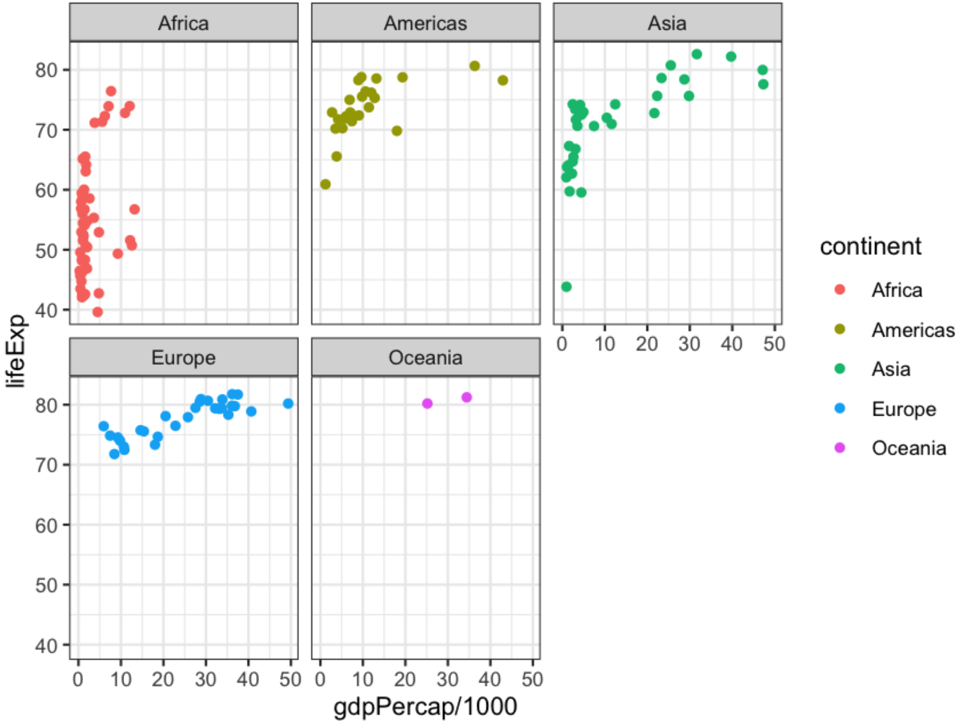
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
                colour = continent)) +
  geom_point() +
  facet_wrap(~continent)
```



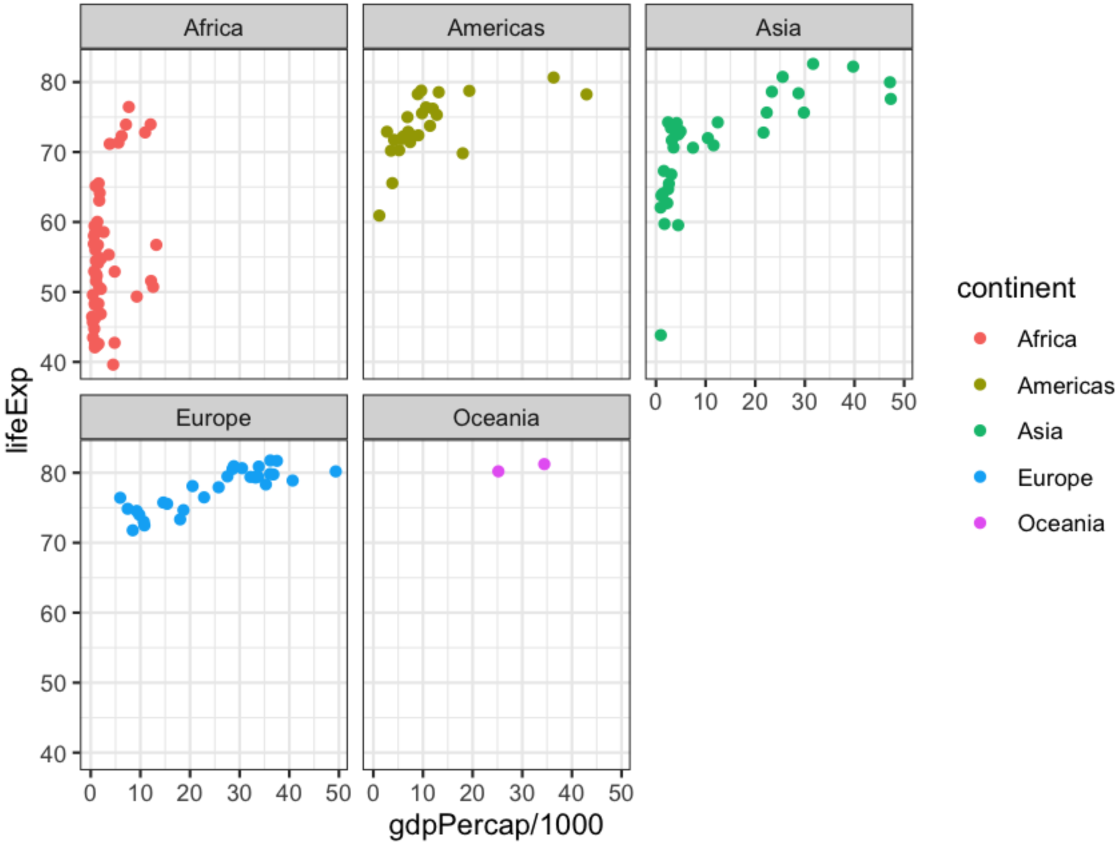
White background - `theme_bw()`

We can also include calculations inside `aes()`: e.g., `x = gdpPercap/1000`

```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap/1000, y = lifeExp,
                colour = continent)) +
  geom_point() +
  facet_wrap(~continent) +
  theme_bw()
```



This is how `ggplot()` works - by adding or modifying things one at a time



Main geoms:

`geom_point()` or `geom_jitter()`

`geom_line()`

`geom_bar()` and `geom_col()`

`geom_histogram()`

`geom_boxplot()`

`geom_label()` or `geom_text()`

These are just the main ones, Google "ggplot gallery" for many more options.

And the `ggplot()` documentation: http://docs.ggplot2.org/

Exercises

https://argoshare.is.ed.ac.uk/ggplot_intro/