ggplot2

HealthyR Bitesize

Artwork by @allison_horst

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

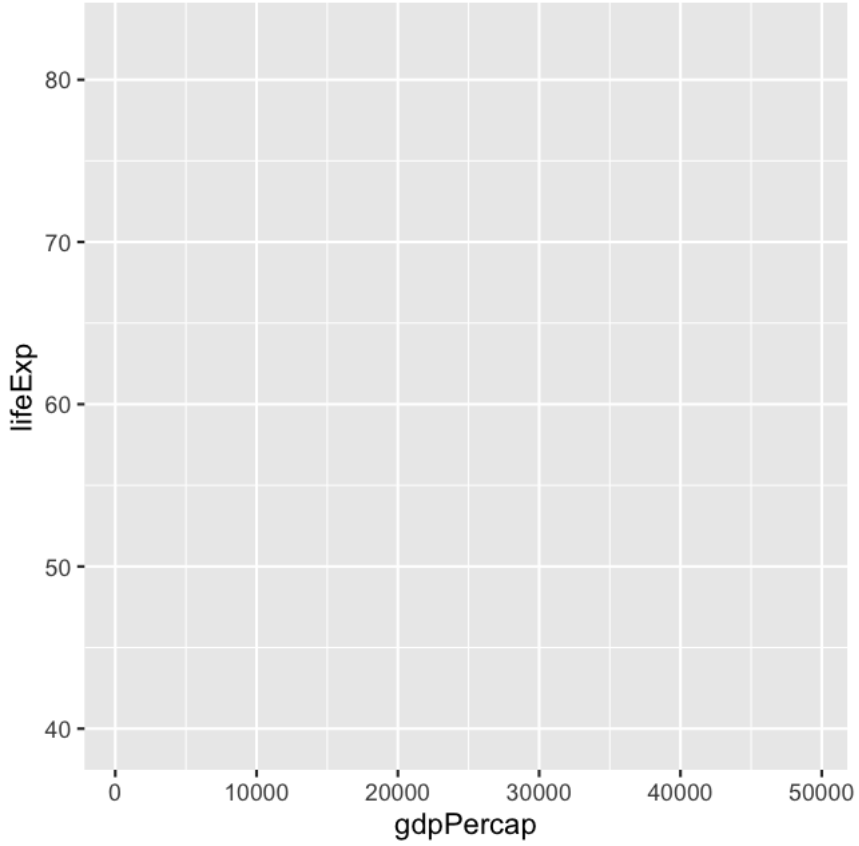
country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1952	28.80	8425333	779.4
Afghanistan	Asia	1957	30.33	9240934	820.9
Afghanistan	Asia	1962	32.00	10267083	853.1
Afghanistan	Asia	1967	34.02	11537966	836.2
Afghanistan	Asia	1972	36.09	13079460	740.0
Afghanistan	Asia	1977	38.44	14880372	786.1
Afghanistan	Asia	1982	39.85	12881816	978.0
Afghanistan	Asia	1987	40.82	13867957	852.4
Afghanistan	Asia	1992	41.67	16317921	649.3
Afghanistan	Asia	1997	41.76	22227415	635.3

The dataset we are using includes 142 countries.

`ggplot()` initialises a canvas

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

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

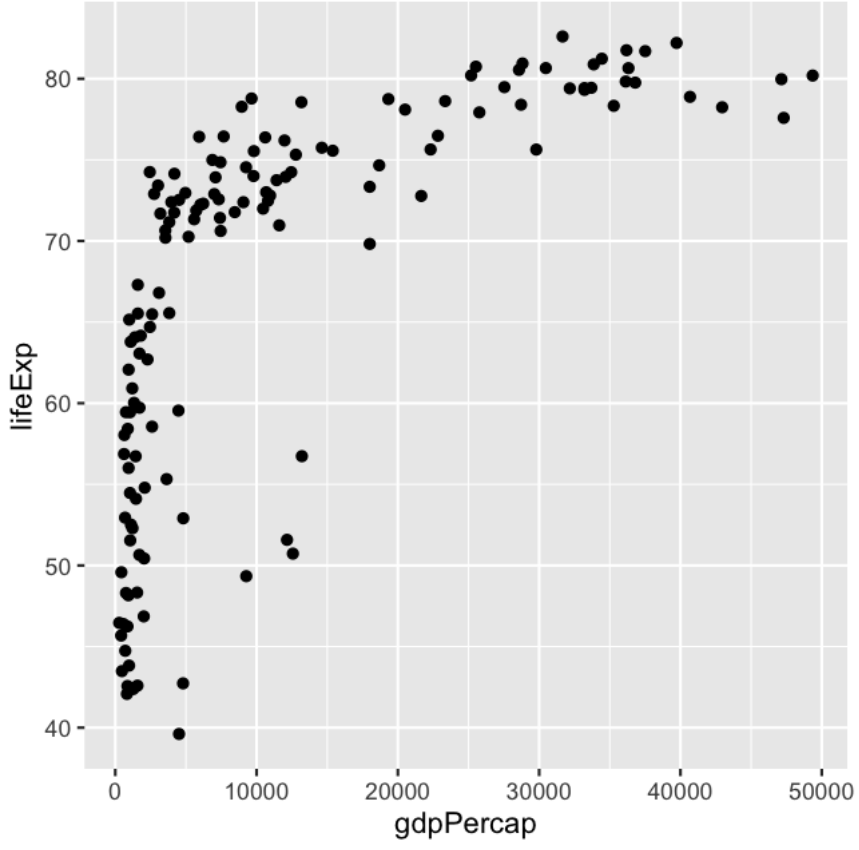


`filter()` is often useful before `ggplot()`

...a canvas to which we can add geoms

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

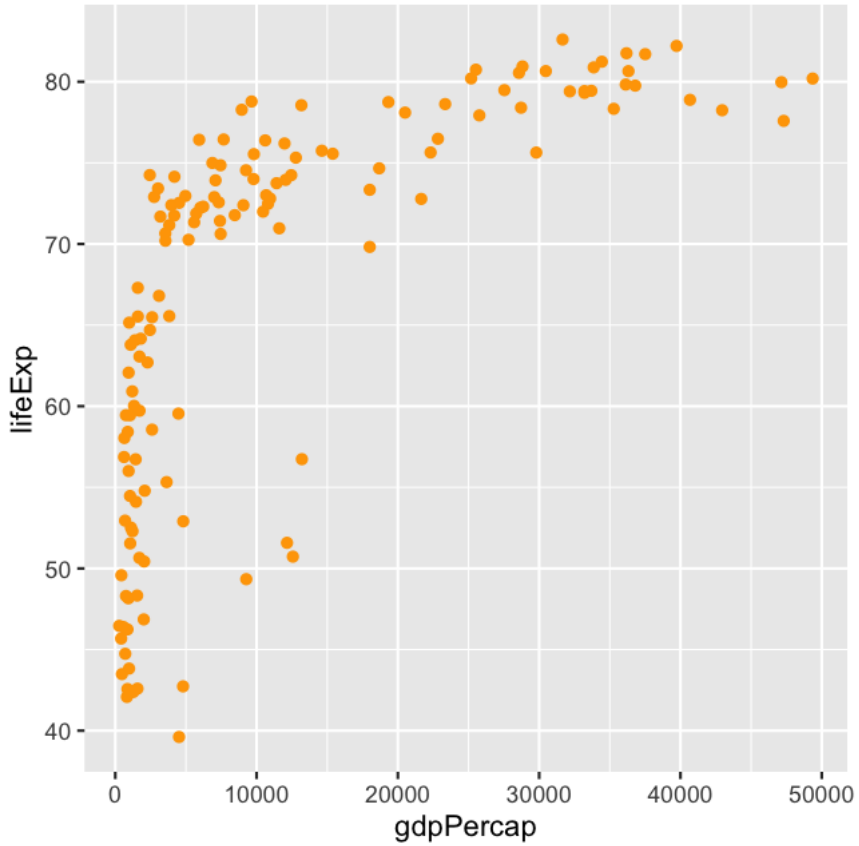
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp)) +
  geom_point()
```



specifying visual properties outside `aes()`

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

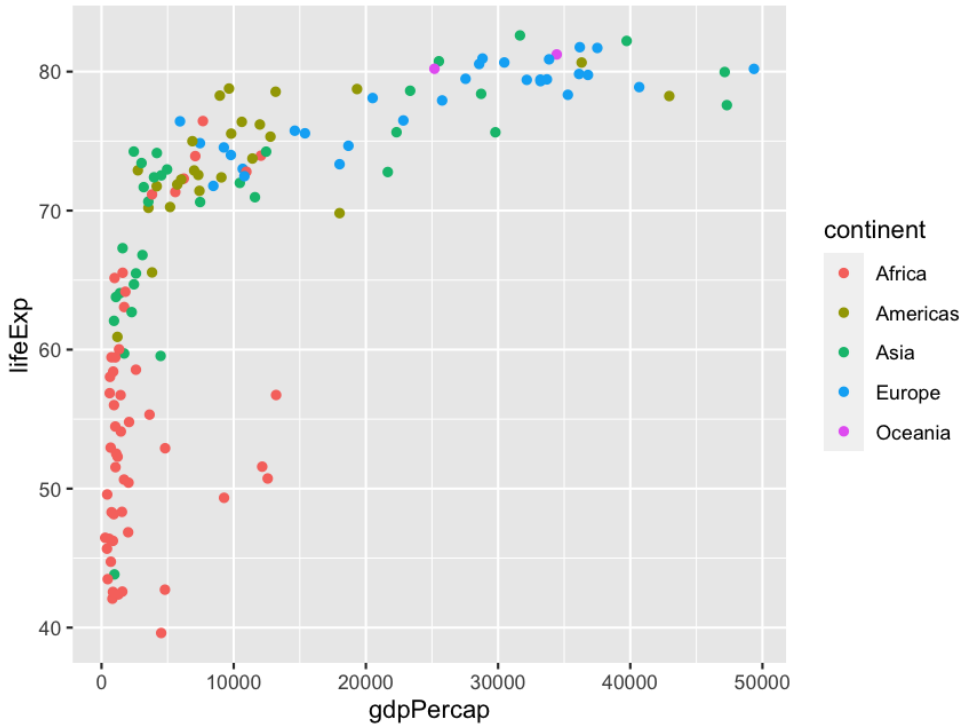
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp)) +
  geom_point(colour = "orange")
```



specifying visual properties inside `aes()`

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

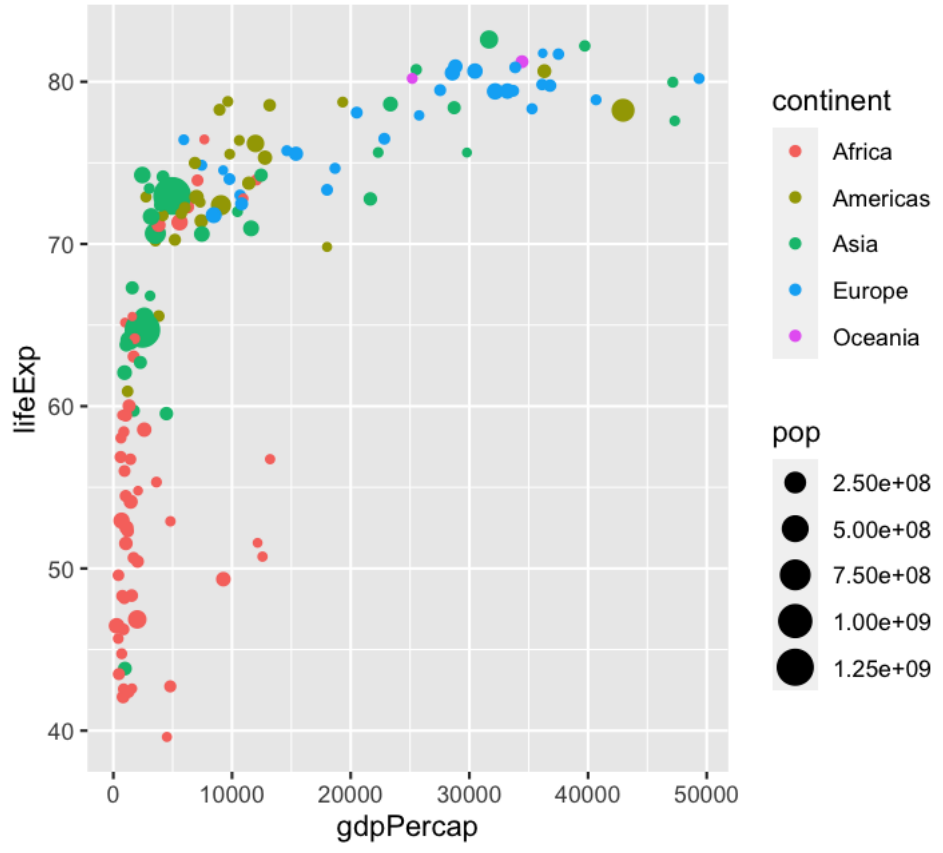
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
             colour = continent)) +
  geom_point()
```



There are many different `aes()` you can include

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

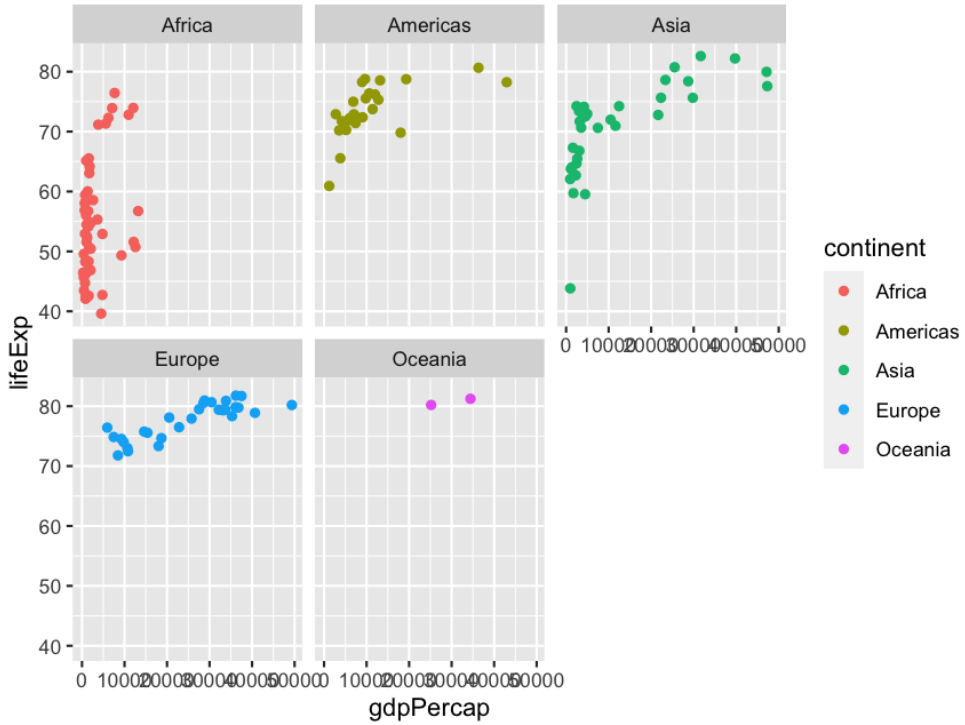
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
             colour = continent, size = pop))
  geom_point()
```



* Press F1 on `geom_point()` to see the list of aesthetics. If F1 is not working, type `geom_point` into the Help tab within RStudio.

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

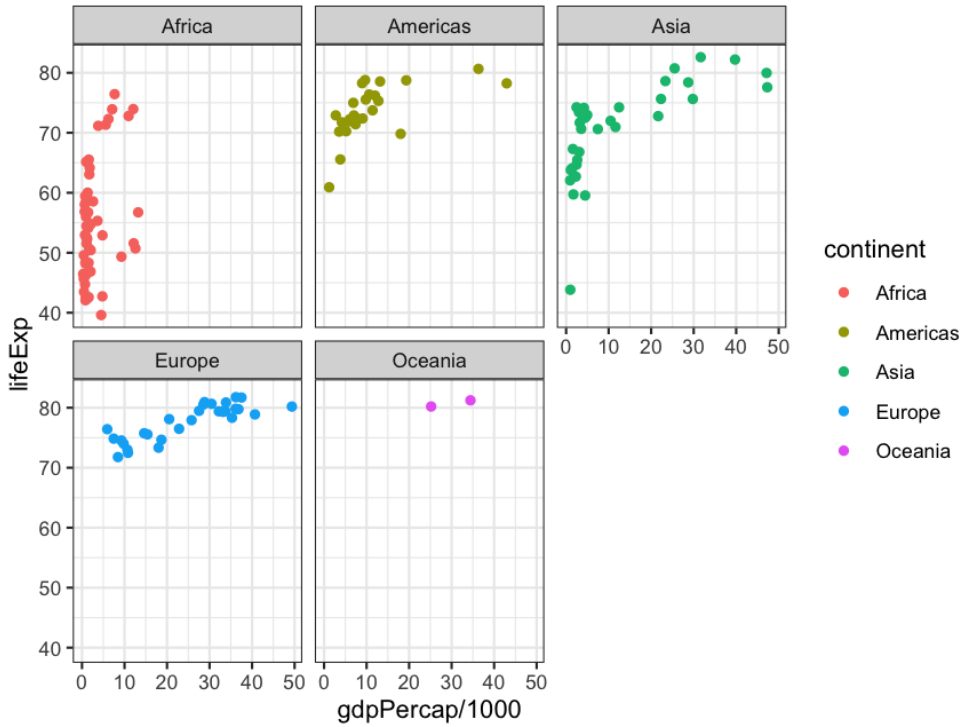
```
gapdata %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
             colour = continent)) +
  geom_point() +
  facet_wrap(~continent)
```



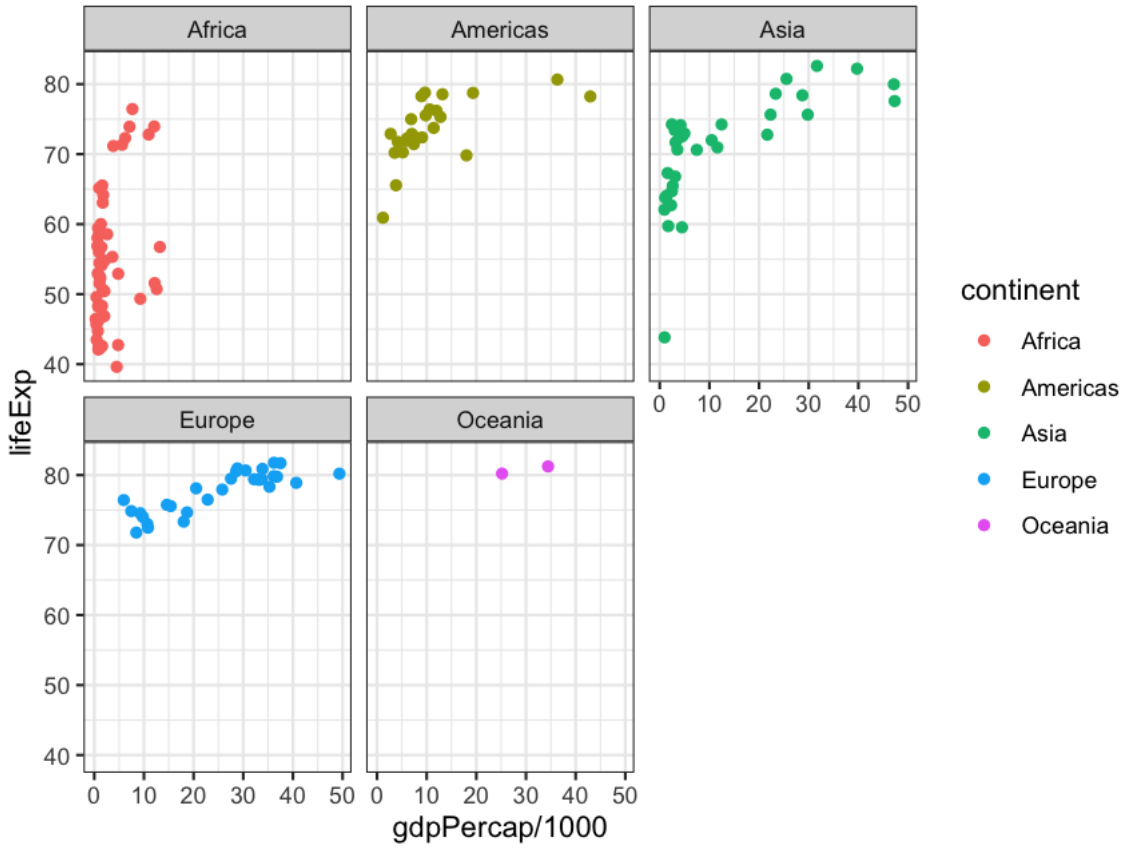
White background - `theme_bw()`

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/

LearnR tutorial

Link to be shared in Zoom Chat window now.