

Class 5: Data Viz with ggplot

Size Size (PID 167)

Quarto

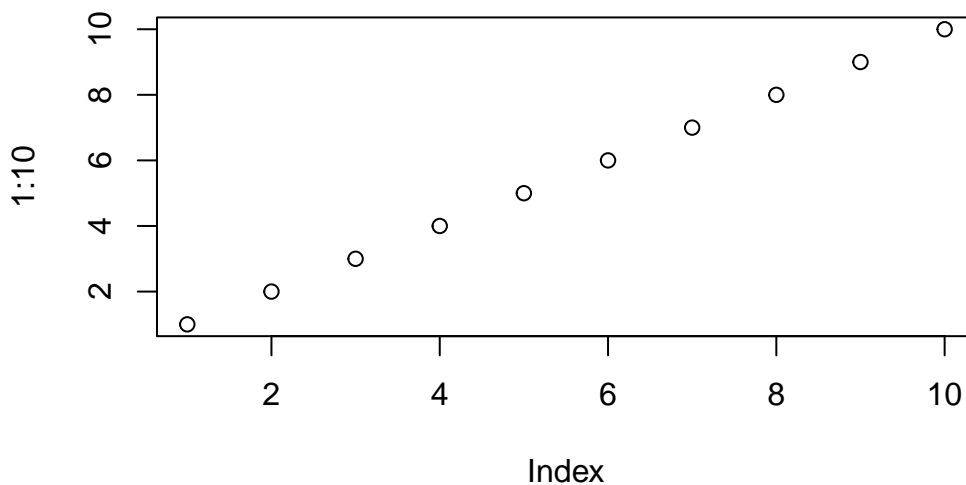
Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

Intro to ggplot

```
plot(1:10)
```



You can add options to executable code like this:

The `echo: false` option disables the printing of code (only output is displayed). First I need to install this package on my computer. To install any R package I use the function `'install.packages()'`

> I will run `'install.packages("ggplot2")` in my R console not this quarto document!

```
#ggplot(cars)
```

Let's see how we can make this figure using **ggplot**

Before I can use any functions from add on packages I need to load the package from my `"library()"` with the `library(ggplot2)` call.

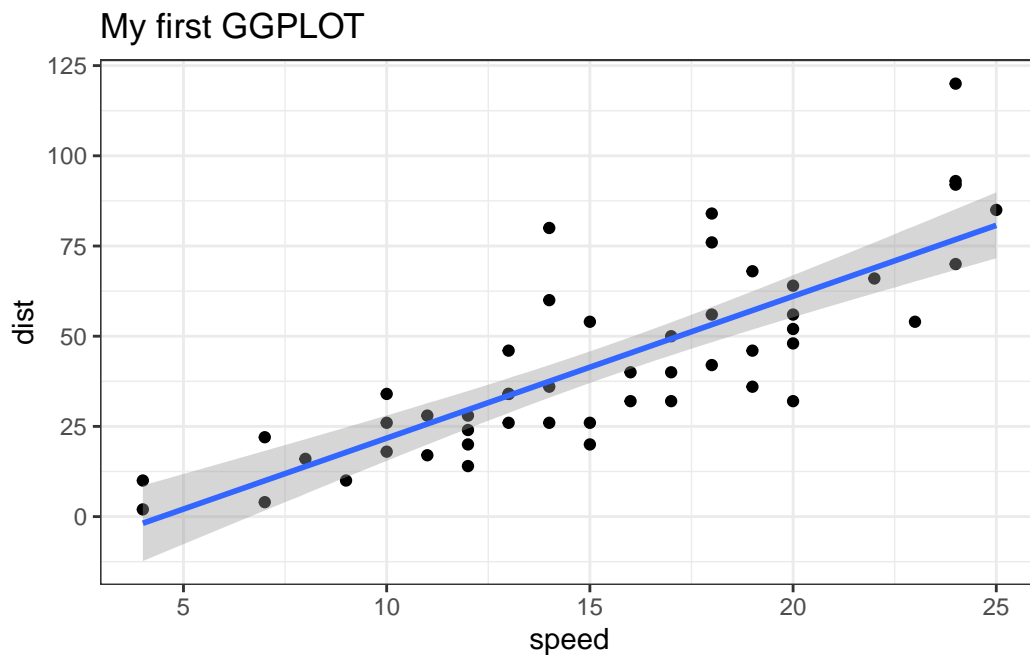
```
library(ggplot2)
```

All ggplot figures have at least 3 things (called layers). These include"

-**data** data (the input dataset I want to plot from) -**aes** (the aesthetic mapping of gthe data to my plot) -**geoms** (the `geom_point()`, `geom_line()` etc. that I want to draw).

```
ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point() +
  geom_smooth(method="lm") +
  theme_bw()+
  labs(title="My first GGPlot")
```

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



```
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

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

down	unchanging	up
72	4997	127

Q. Using your values above and 2 significant figures. What fraction of total genes is up regulated in this dataset?

```
round(table(genes$State)/nrow(genes), 4)
```

down	unchanging	up
0.0139	0.9617	0.0244

```
n.tot <- nrow(genes)
vals <- table(genes$state)

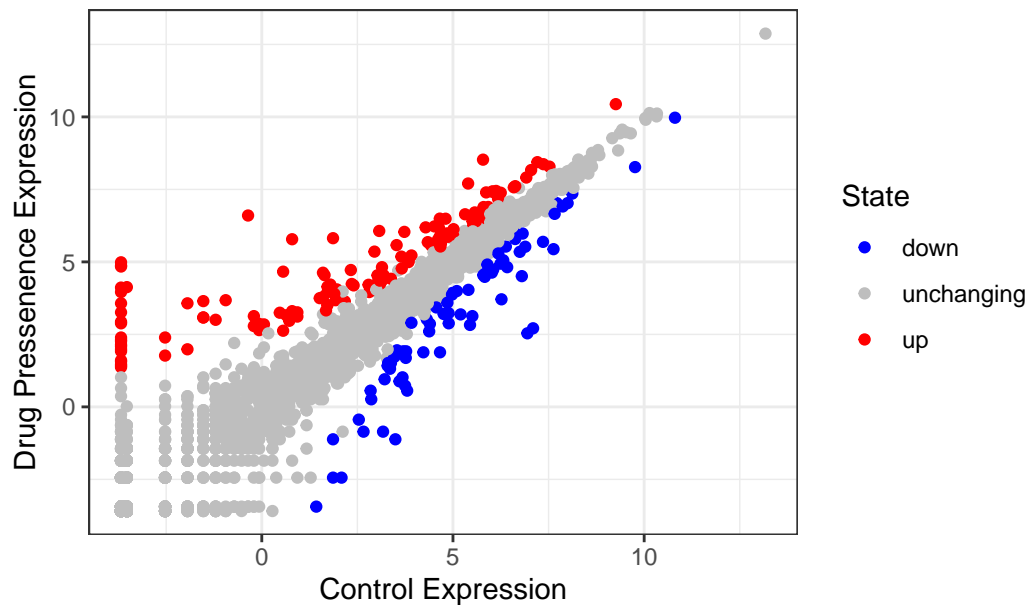
vals.percent <- vals/n.tot * 100
round(vals.percent, 2)
```

```
numeric(0)
```

A first plot of this dataset

```
ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point() +
  theme_bw() +
  labs(title="Gene Expression changes upon drug treatment",
       x="Control Expression",
       y= "Drug Pressence Expression") +
  scale_color_manual( values=c("blue","gray","red") )
```

Gene Expression changes upon drug treatment



```
#read in gapminder file
```

```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"
gapminder <- read.delim(url)
```

```
#call dplyr library and filter for year 2007
```

```
# install.packages("dplyr") ## un-comment to install if needed
library(dplyr)
```

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':
```

```
filter, lag
```

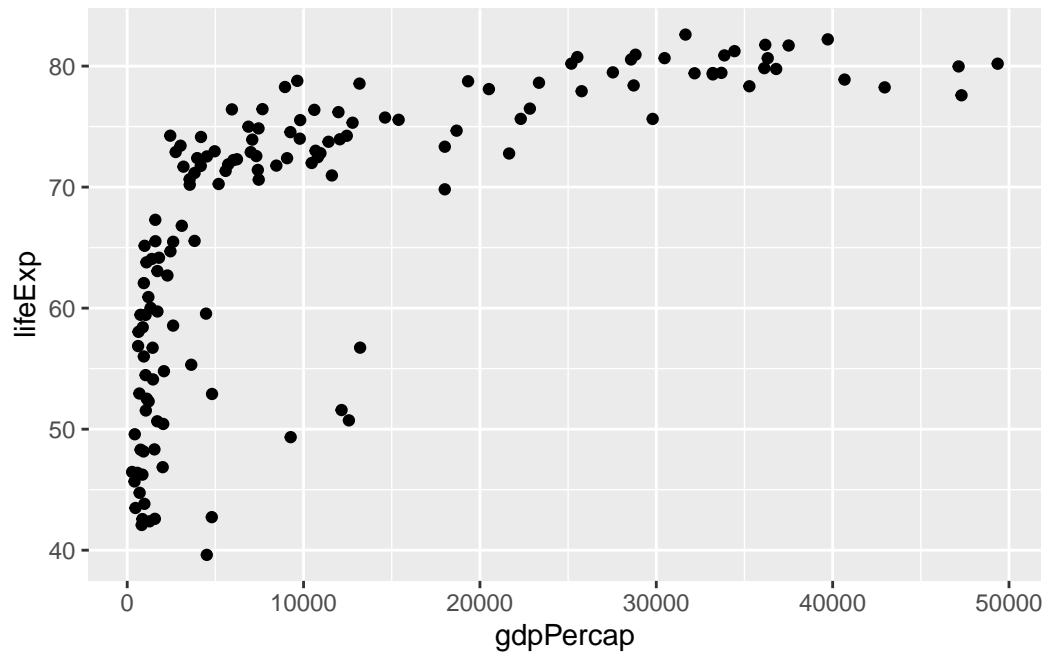
```
The following objects are masked from 'package:base':
```

```
intersect, setdiff, setequal, union
```

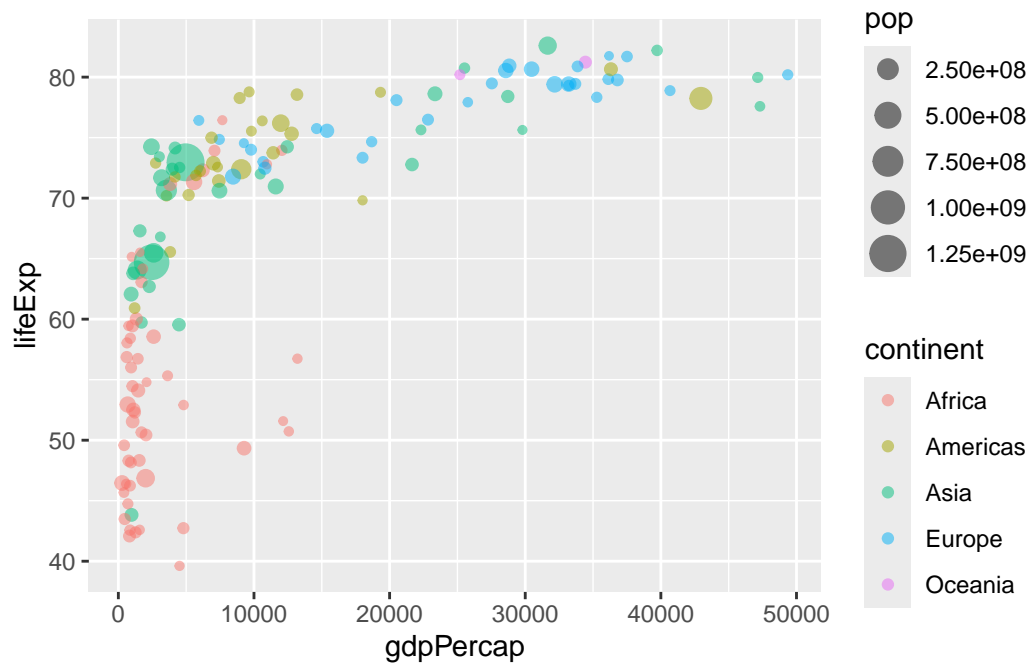
```
gapminder_2007 <- gapminder %>% filter(year==2007)
```

```
#plot
```

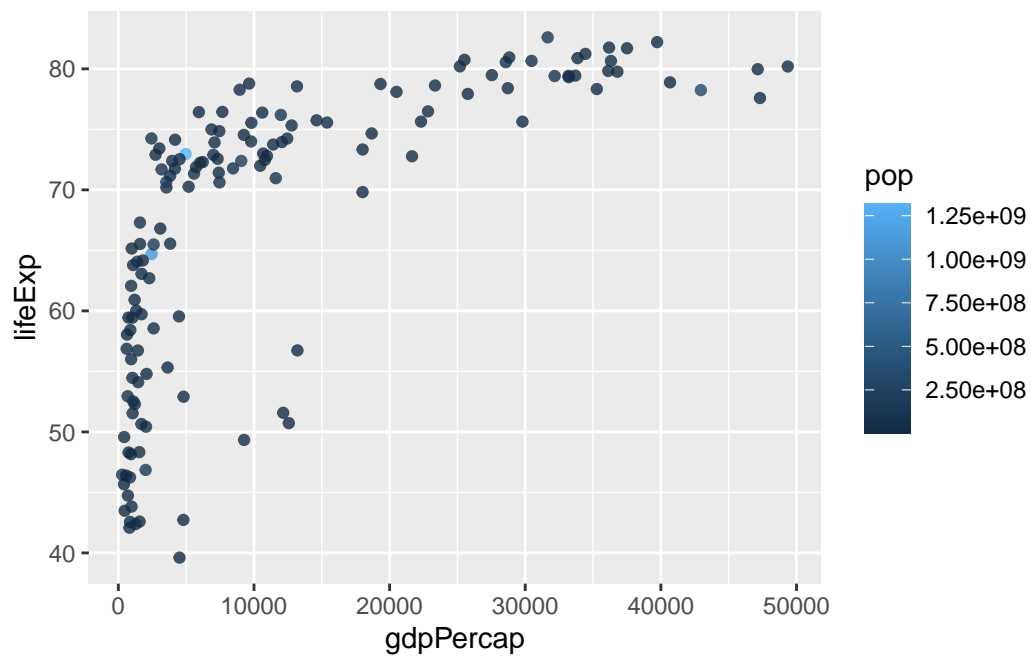
```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp) +  
  geom_point()
```



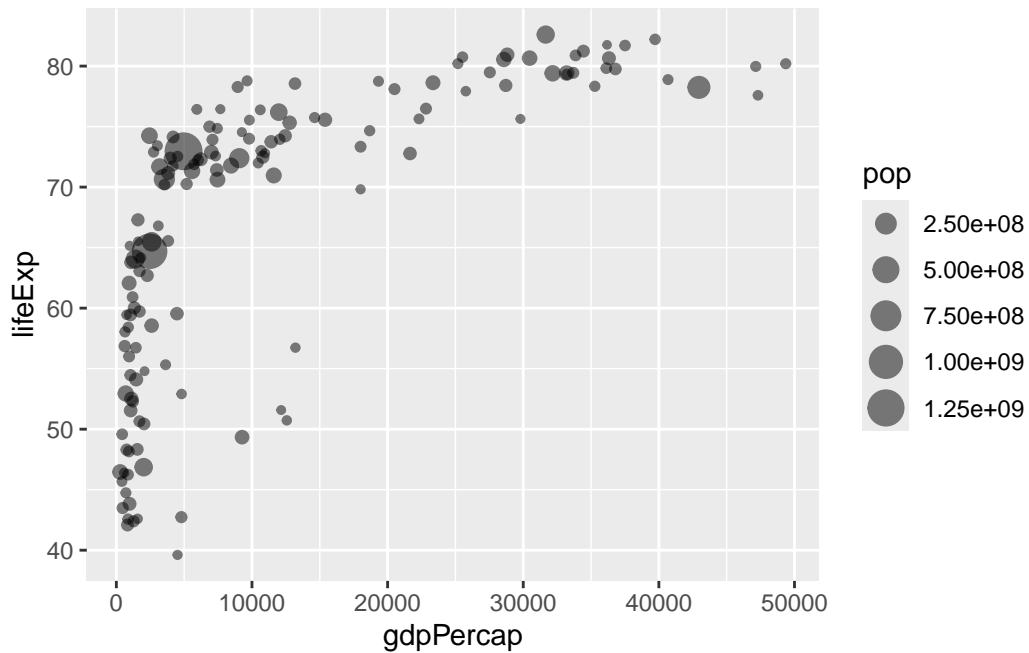
```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +  
  geom_point(alpha=0.5)
```



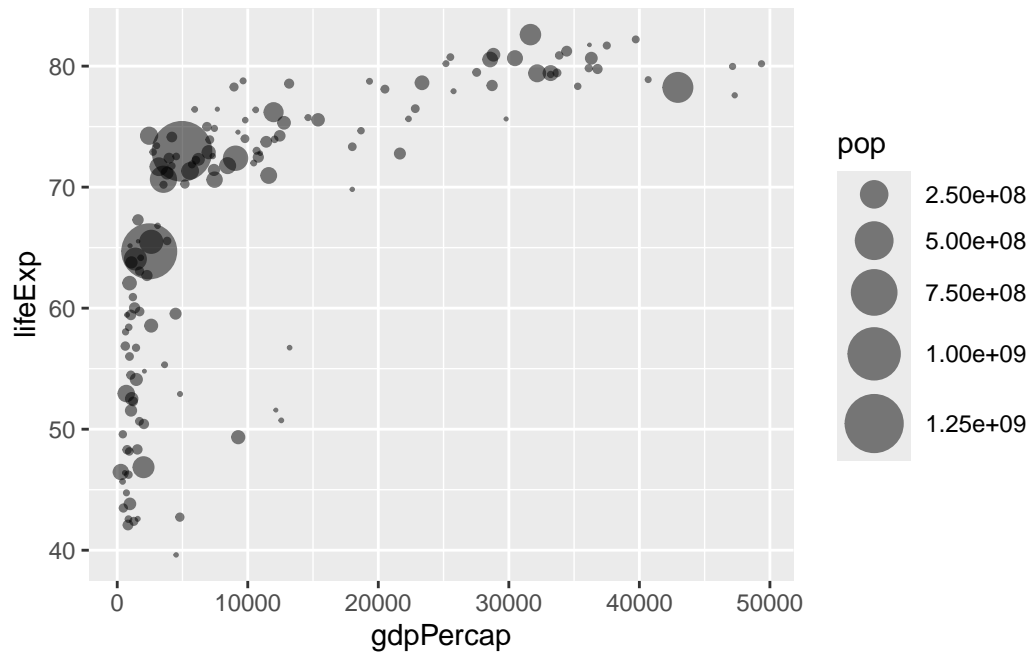
```
ggplot(gapminder_2007) +  
  aes(x = gdpPerCap, y = lifeExp, color = pop) +  
  geom_point(alpha=0.8)
```



```
ggplot(gapminder_2007) +
  aes(x = gdpPerCap, y = lifeExp, size = pop) +
  geom_point(alpha=0.5)
```

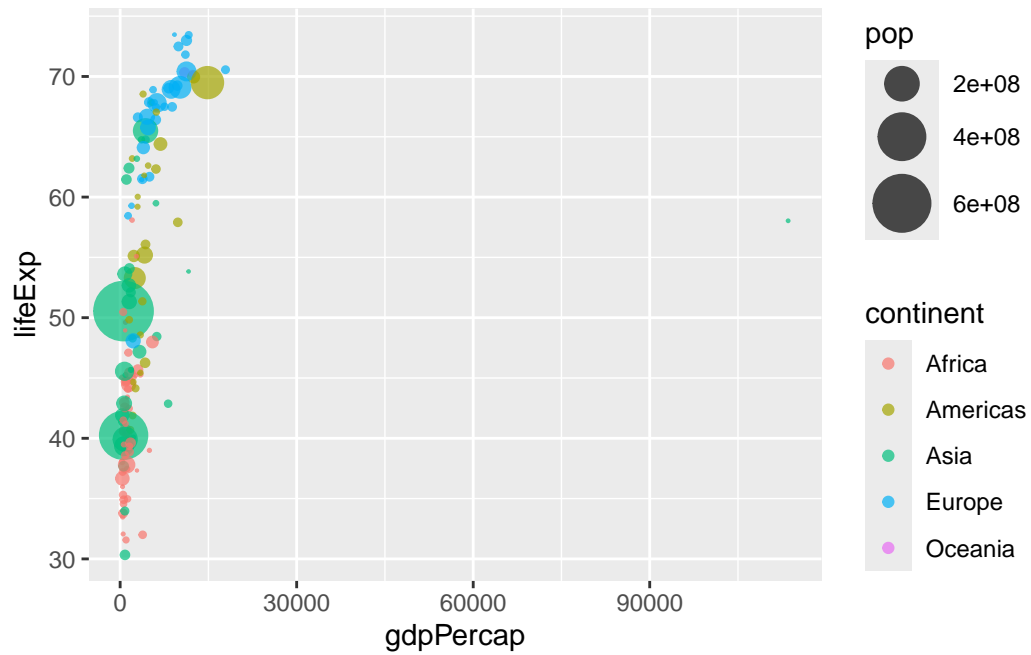


```
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPerCap, y = lifeExp,
                 size = pop), alpha=0.5) +
  scale_size_area(max_size = 10)
```

```
gapminder_1957 <- gapminder %>% filter(year==1957)

ggplot(gapminder_1957) +
  aes(x = gdpPercap, y = lifeExp, color=continent,
      size = pop) +
  geom_point(alpha=0.7) +
  scale_size_area(max_size = 10)
```



```
gapminder_1957 <- gapminder %>% filter(year==1957 | year==2007)

ggplot(gapminder_1957) +
  geom_point(aes(x = gdpPerCap, y = lifeExp, color=continent,
                 size = pop), alpha=0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```

