Annotating a plot

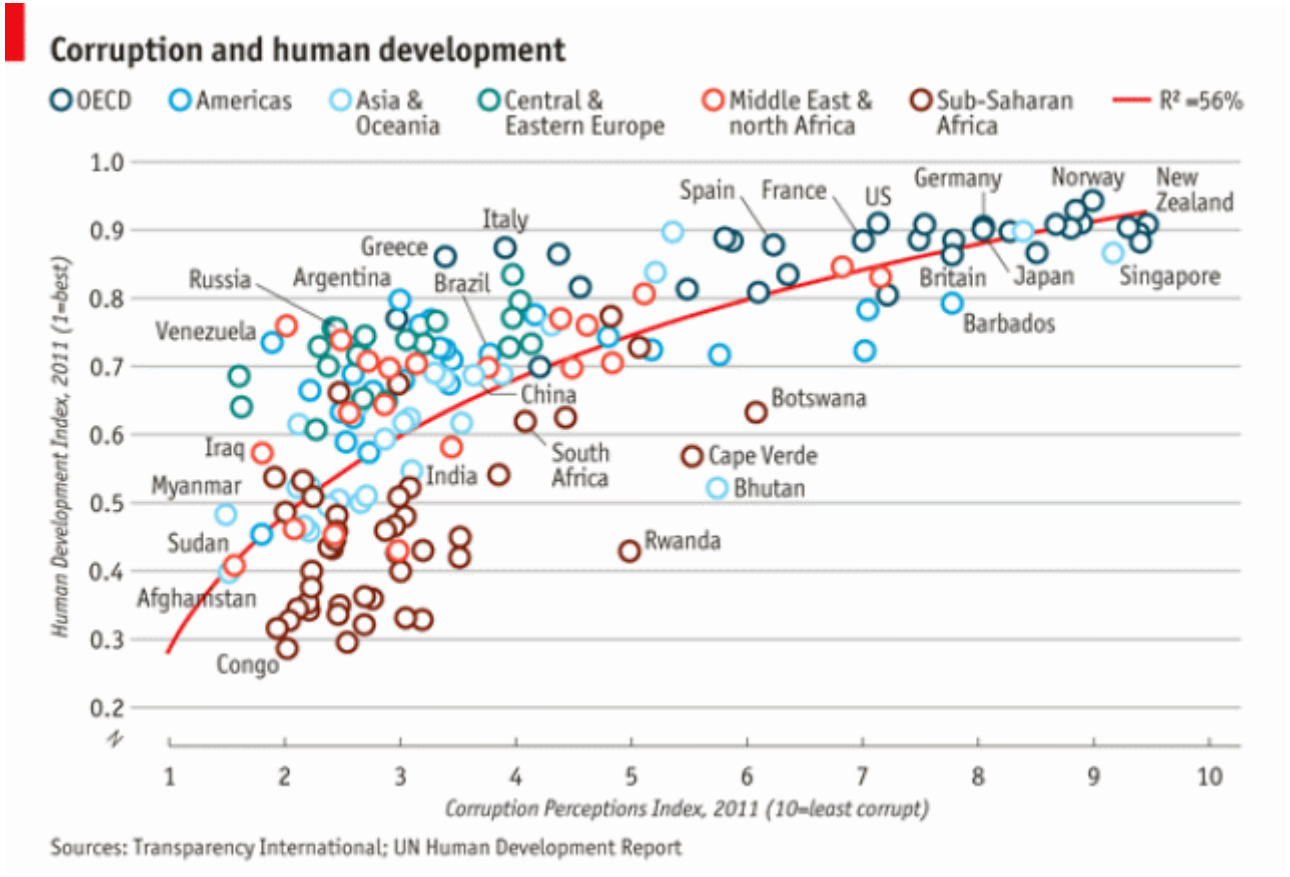
Abhijit Dasgupta, PhD

Annotations

Stand-alone stories

- You would like a data visualization to stand on its own
- Relevant information should be placed on the graph
- However, you need to balance the information content with real estate
    - Don't clutter the graph and make it not readable

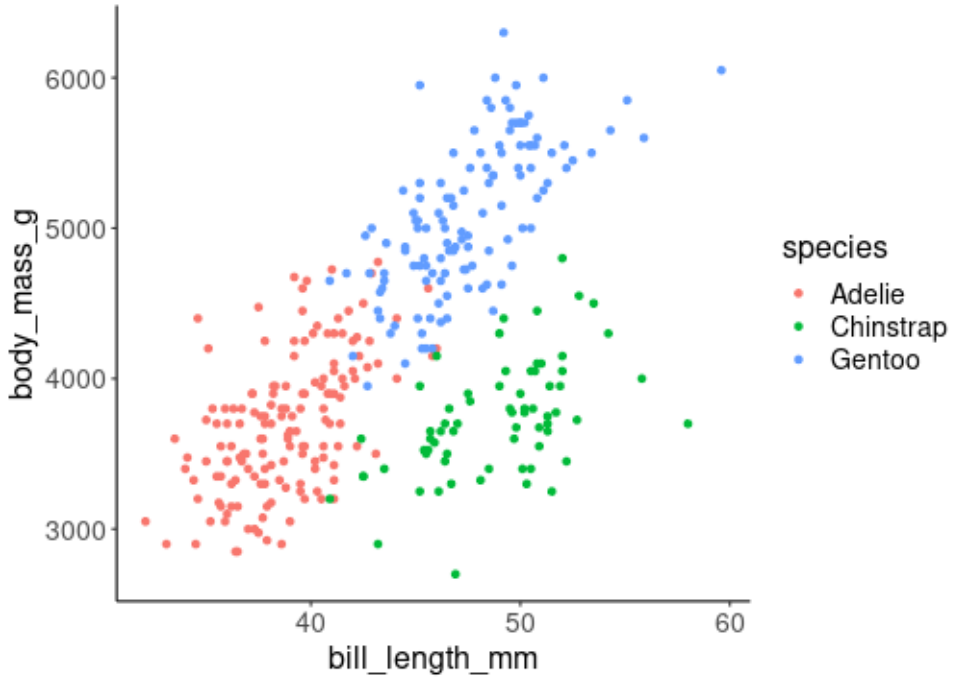
An example



We will recreate this plot in a tutorial

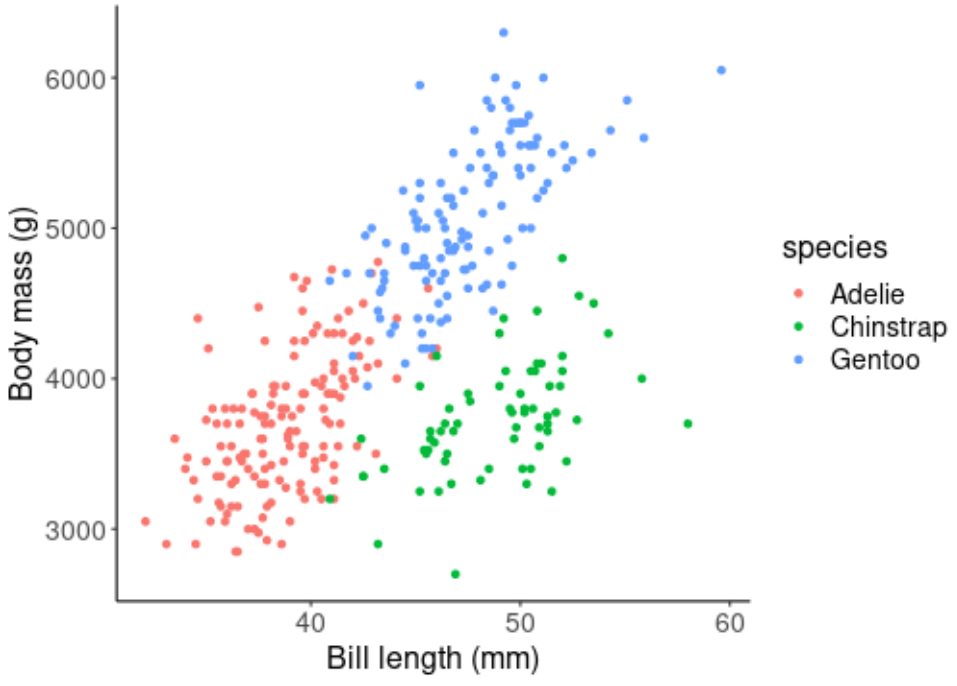
Titles, subtitles and captions

```
library(palmerpenguins)
(plt <- ggplot(penguins,
               aes(bill_length_mm, body_mass_g,
                   color=species))+
  geom_point())
```



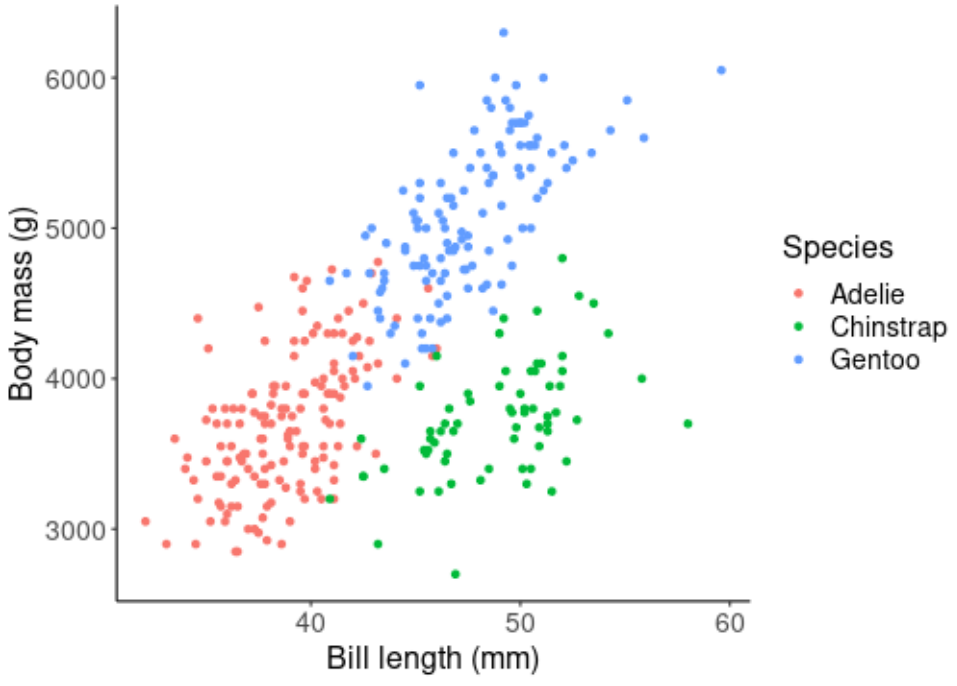
Titles, subtitles and captions

```
library(palmerpenguins)
plt <- ggplot(penguins,
              aes(bill_length_mm, body_mass_g,
                  color=species))+
  geom_point()
plt +
  labs(x = 'Bill length (mm)',
       y = 'Body mass (g)')
```



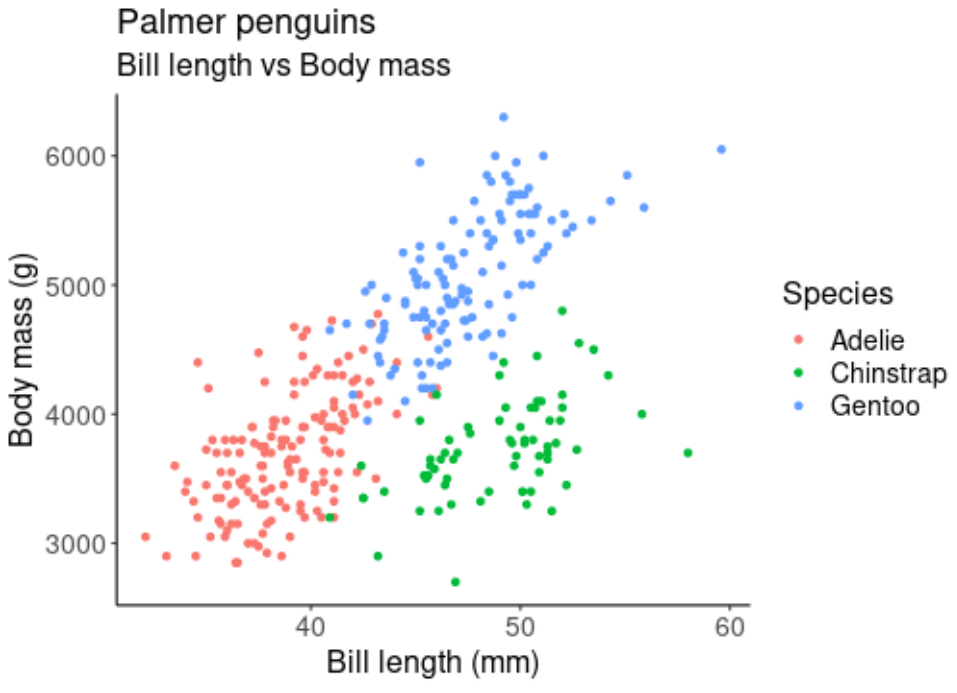
Titles, subtitles and captions

```
library(palmerpenguins)
plt <- ggplot(penguins,
              aes(bill_length_mm, body_mass_g,
                  color=species))+
  geom_point()
plt +
  labs(x = 'Bill length (mm)',
       y = 'Body mass (g)',
       color = 'Species')
```



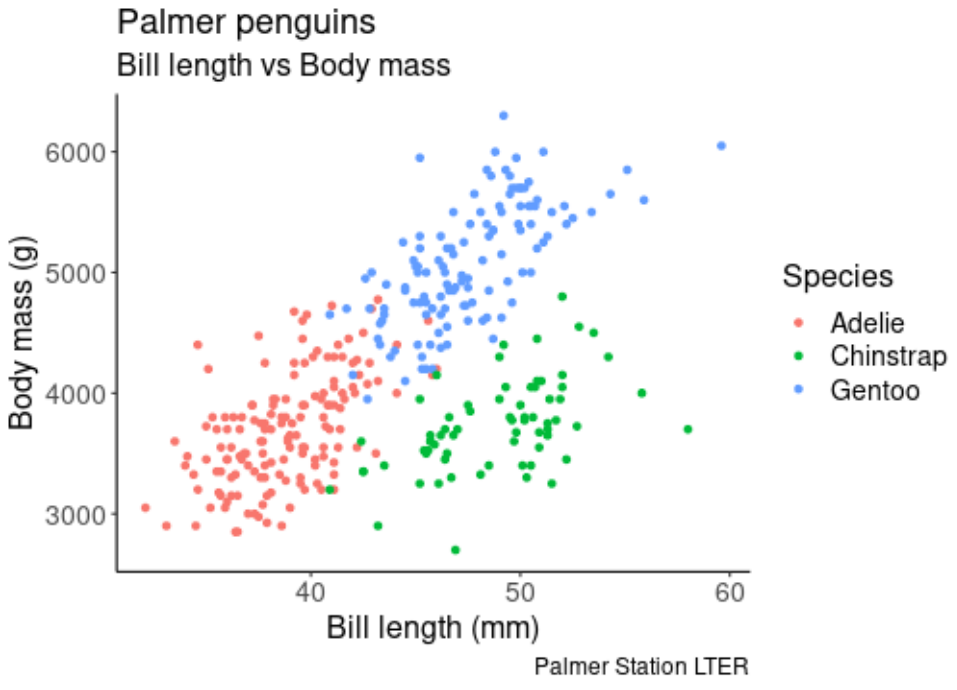
Titles, subtitles and captions

```
library(palmerpenguins)
plt <- ggplot(penguins,
              aes(bill_length_mm, body_mass_g,
                  color=species))+
  geom_point()
plt +
  labs(x = 'Bill length (mm)',
       y = 'Body mass (g)',
       color = 'Species',
       title = "Palmer penguins",
       subtitle = "Bill length vs Body mass")
```



Titles, subtitles and captions

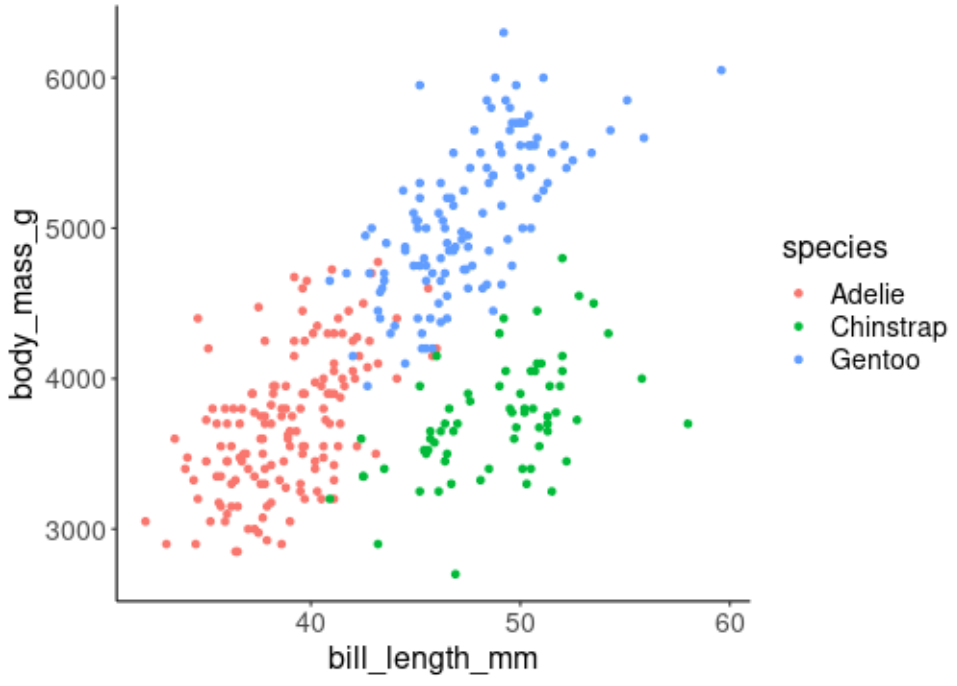
```
library(palmerpenguins)
plt <- ggplot(penguins,
              aes(bill_length_mm, body_mass_g,
                  color=species))+
  geom_point()
plt +
  labs(x = 'Bill length (mm)',
       y = 'Body mass (g)',
       color = 'Species',
       title = "Palmer penguins",
       subtitle = "Bill length vs Body mass",
       caption = "Palmer Station LTER")
```



Adding derived statistics to a plot

Adding group means

```
ggplot(penguins,
       aes(x = bill_length_mm,
           y = body_mass_g,
           color = species))+
  geom_point()
```

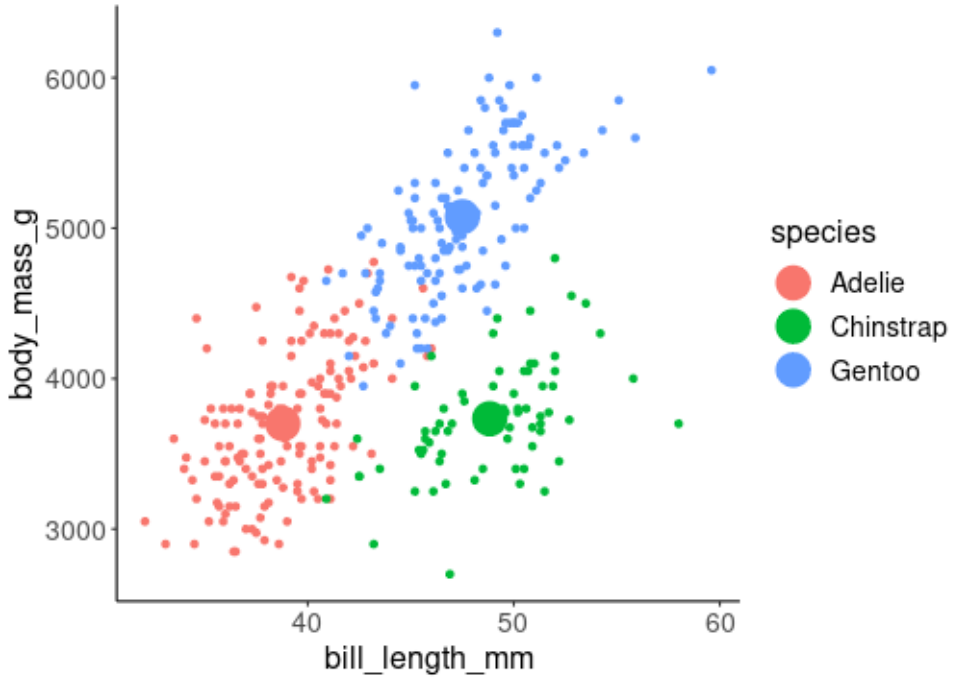


Adding group means

```
means <- penguins %>% group_by(species) %>%
  summarize_at(vars(bill_length_mm, body_mass_g),
               mean, na.rm=TRUE)
means
```

```
  # A tibble: 3 x 3
    species   bill_length_mm body_mass_g
    <fct>              <dbl>       <dbl>
  1 Adelie              38.8       3701.
  2 Chinstrap           48.8       3733.
  3 Gentoo              47.5       5076.
```

```
ggplot(penguins,
       aes(x = bill_length_mm,
           y = body_mass_g,
           color = species))+
  geom_point()+
  geom_point(data = means,
             size=8)
```



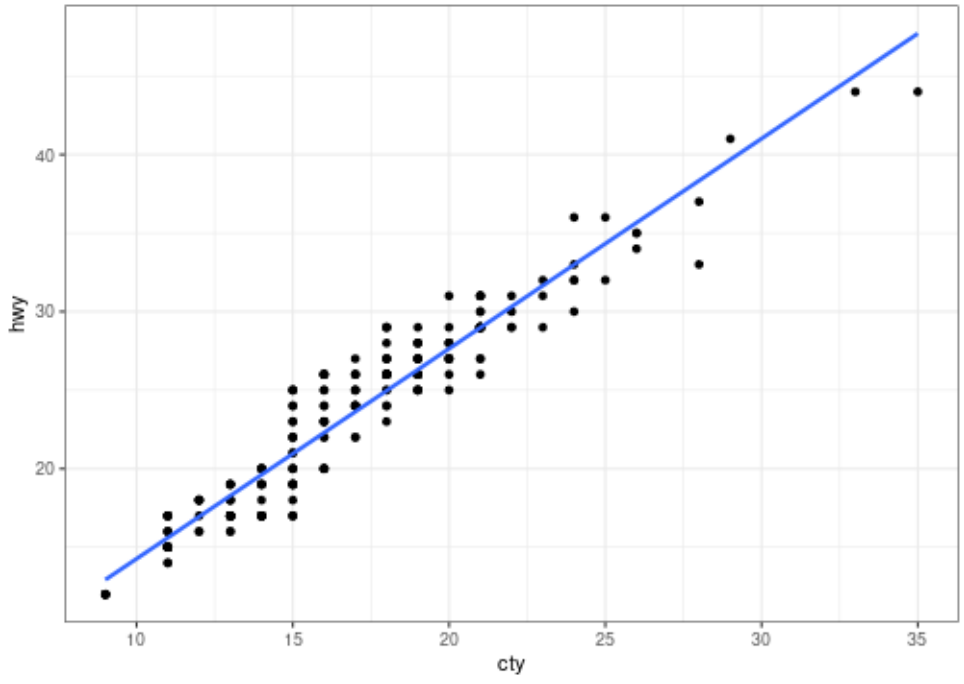
Adding data from a different dataset

Adding regression metrics

Regress highway mileage on city mileage (data: mpg)

```
mod1 <- lm(hwy ~ cty, data = mpg)
r2 <- broom::glance(mod1) %>% pull(r.squared)

ggplot(mpg,
       aes(x = cty, y = hwy))+
  geom_point() +
  geom_smooth(method = 'lm', se=F) +
  theme_bw()
```

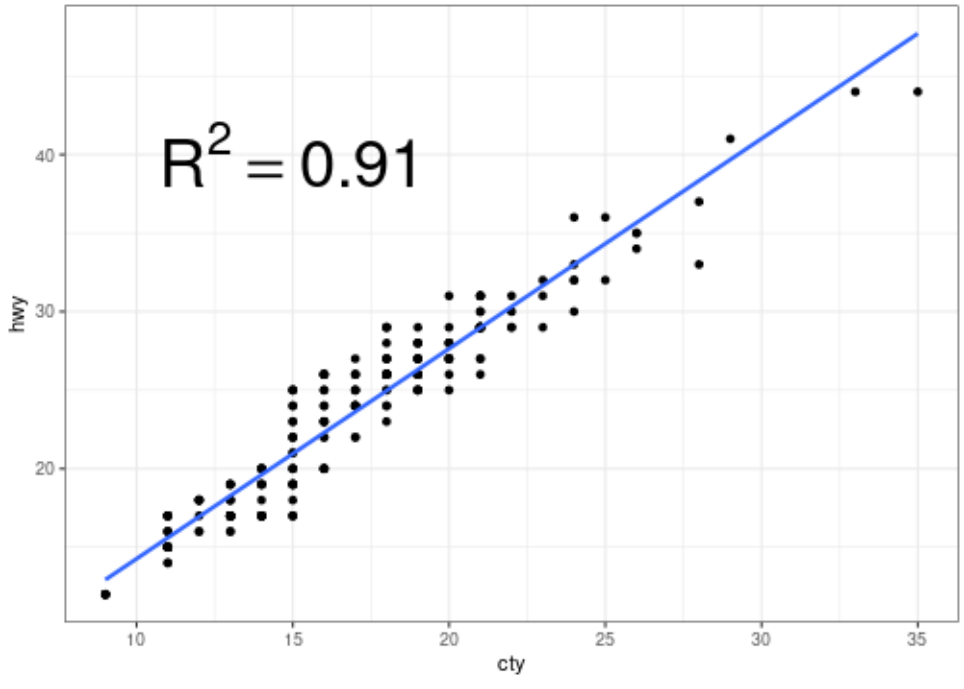


Adding regression metrics

Regress highway mileage on city mileage (data: mpg)

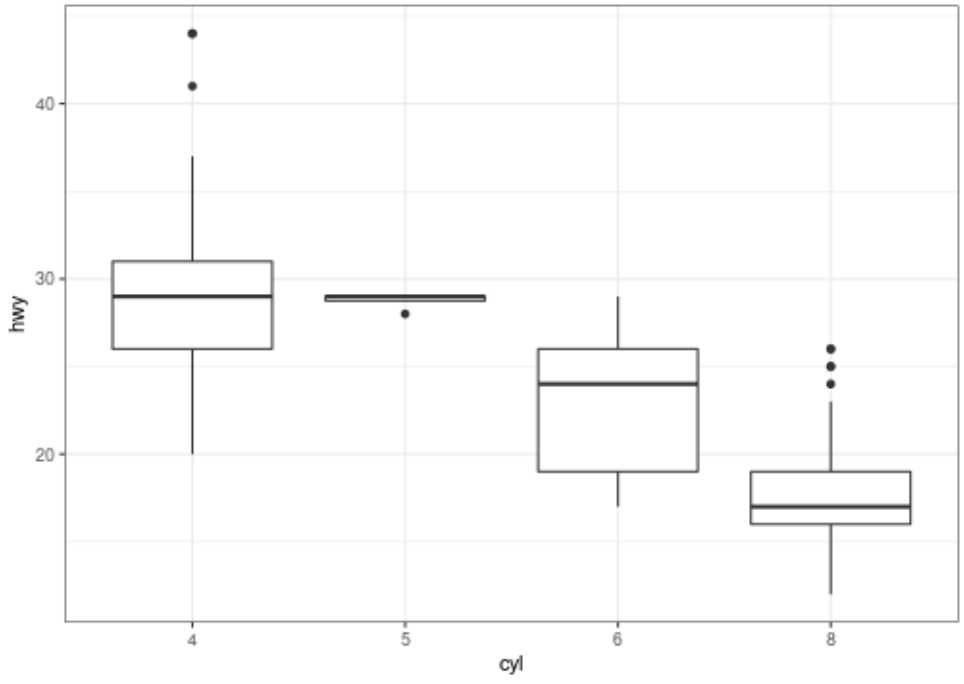
```
mod1 <- lm(hwy ~ cty, data = mpg)
r2 <- broom::glance(mod1) %>% pull(r.squared) %>%
  round(., 2)

ggplot(mpg,
       aes(x = cty, y = hwy))+
  geom_point() +
  geom_smooth(method = 'lm', se=F)+
  annotate(geom='text',
           x = 15, y = 40,
           label=glue::glue("R^2 == {r}",r=r2),
           size=12,
           parse=T) +
  theme_bw()
```



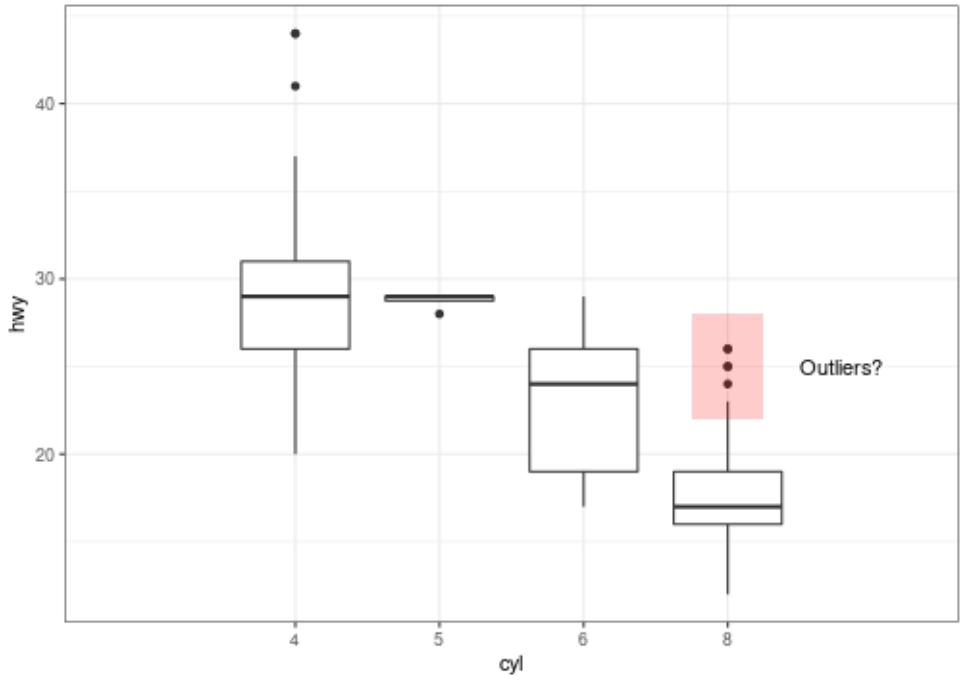
Highlighting regions

```
mpg %>%
  mutate(cyl = as.factor(cyl)) %>%
  ggplot(aes(x = cyl, y = hwy)) +
  geom_boxplot() +
  theme_bw()
```



Highlighting regions

```
mpg %>%
  mutate(cyl = as.factor(cyl)) %>%
  ggplot(aes(x = cyl, y = hwy)) +
  geom_boxplot() +
  theme_bw()+
  annotate(geom = 'rect',
           xmin=3.75,xmax=4.25,
           ymin = 22, ymax = 28,
           fill = 'red',
           alpha = 0.2) +
  annotate('text',
           x = 4.5, y = 25,
           label = 'Outliers?',
           hjust = 0)+
  coord_cartesian(xlim = c(0,5))+
  theme_bw()
```



Note: If you have a factor on the x-axis, they are plotted at 1, 2, 3, ...