Exercises: ggplot2

Nathan Green, Imperial College London

07/09/2019

First read in the data

```
library(dplyr)
library(ggplot2)
library(dataPakistan)

file_location <- system.file("extdata", package = "dataPakistan")
dat <- readxl::read_excel(path = paste0(file_location, "/Admin-datasheet-year2018.xlsx"))
```

1. Use `qplot()` to plot basic scatter plot of `Coverage at Hosehold (n)` vs `Coverage at Hosehold (%)`.
2. Most of the points are squashed to the left. Use `xlim()` to zoom in on to these.
3. Use `ggplot()` to create the same plot.
4. Assign the ggplot component to a variable and create the same plot as above.
5. Using

```
install.packages("hexbin")
library(hexbin)
```

Re-plot as a hexplot with 50 bins.

6. Plot the scatterplot with opaque points.
7. Make the points a different colour.
8. Colour the points according to `Province`.
9. Colour the points according to `Year` and plot `Province` by `Coverage at Hosehold (n)`. What happens when the value is continuous? How can you fix this? Is there a more appropriate type of plot?
10. Re-do the above as a boxplot.
11. Plot the data points as well using jitter and change the colour.
12. The boxplot his obscured by the points. How do you make the boxes more prominent?

Time series

1. Plot the `Coverage at Hosehold (n)` over `Year` as a line plot.
2. This doesnt distinguish areas, so only plot `ATTOCK` by subsetting the data.
3. Colour the lines by `District`.

Faceting

Only use the first 9 District by running the following

```
nms <- table(dat$District) %>% names %>% .[1:9]
ss_dat <- dat[dat$District %in% nms, ]
```

1. Plot `Year` by `Coverage at Hosehold (n)` and `facet_wrap` by `District` to get a grid of plots

2. Change the colour theme to black and white.

Arranging and exporting plots

Install and load this package

```
install.packages("gridExtra")
library(gridExtra)
```

Define

```
p1 <-
  ggplot(data = ss_dat, mapping = aes(x = Year, y = `Coverage at Hosehold (n)`)) +
  geom_line() +
  facet_wrap(facets = vars(District))

p2 <-
  ggplot(data = ss_dat, mapping = aes(x = Year, y = `Coverage at Hosehold (n)`)) +
  geom_line() +
  facet_wrap(facets = vars(District)) +
  theme_bw()
```

1. Use `grid.arrange()` to make a 2 column figure.
2. Save as a ggplot with dpi of 300.