Pokemon GO data analysis

Levi Waldron

- Load the Pokemon dataset
- Linear models on Combat Power
- `dplyr` introduction
- `ggplot2` introduction
    - A blank canvas
    - A scatterplot
    - A scatterplot plus regression lines
    - Boxplot

Load the Pokemon dataset

Pokemon dataset from https://www.openintro.org/stat/data/?data=pokemon (https://www.openintro.org/stat/data/?data=pokemon).

Option 1, after downloading the csv file to the same directory as this script (PokemonGO.Rmd):

```
pokemon <- read.csv("pokemon.csv")
```

Option 2, downloading from the internet:

```
pokemon = read.csv("https://www.openintro.org/stat/data/pokemon.csv")
```

Linear models on Combat Power

```
pfit = lm(cp ~ species + hp + weight + height, data=pokemon)
pfit2 = lm(cp ~ (species + hp + weight + height)^2, data=pokemon) #all 2-way interaction
s
pfit3 = lm(cp ~ . - notes - name - attack_weak, data=pokemon)  # all available predictor
s except for notes, names, and attack_weak.
```

`dplyr` introduction

Help at https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf (https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf)

Create a dataset that excludes all columns whose name ends in “new” (notice the - before `ends_with()`:

```
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
```

```
pokemon2 = select(pokemon, -ends_with("new"))
```

Exclude columns whose name ends in new, then keep only rows where the species is “Pidgey”:

```
pokemon2 = pokemon %>%
  select(-ends_with("new")) %>%
  filter(species == "Pidgey")
summary(pokemon$species)
```

```
## Caterpie    Eevee   Pidgey   Weedle
##       10        6       39       20
```

```
summary(pokemon2$species)
```

```
## Caterpie    Eevee   Pidgey   Weedle
##        0        0       39        0
```

`ggplot2` introduction

Help at https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf (https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf)

A blank canvas

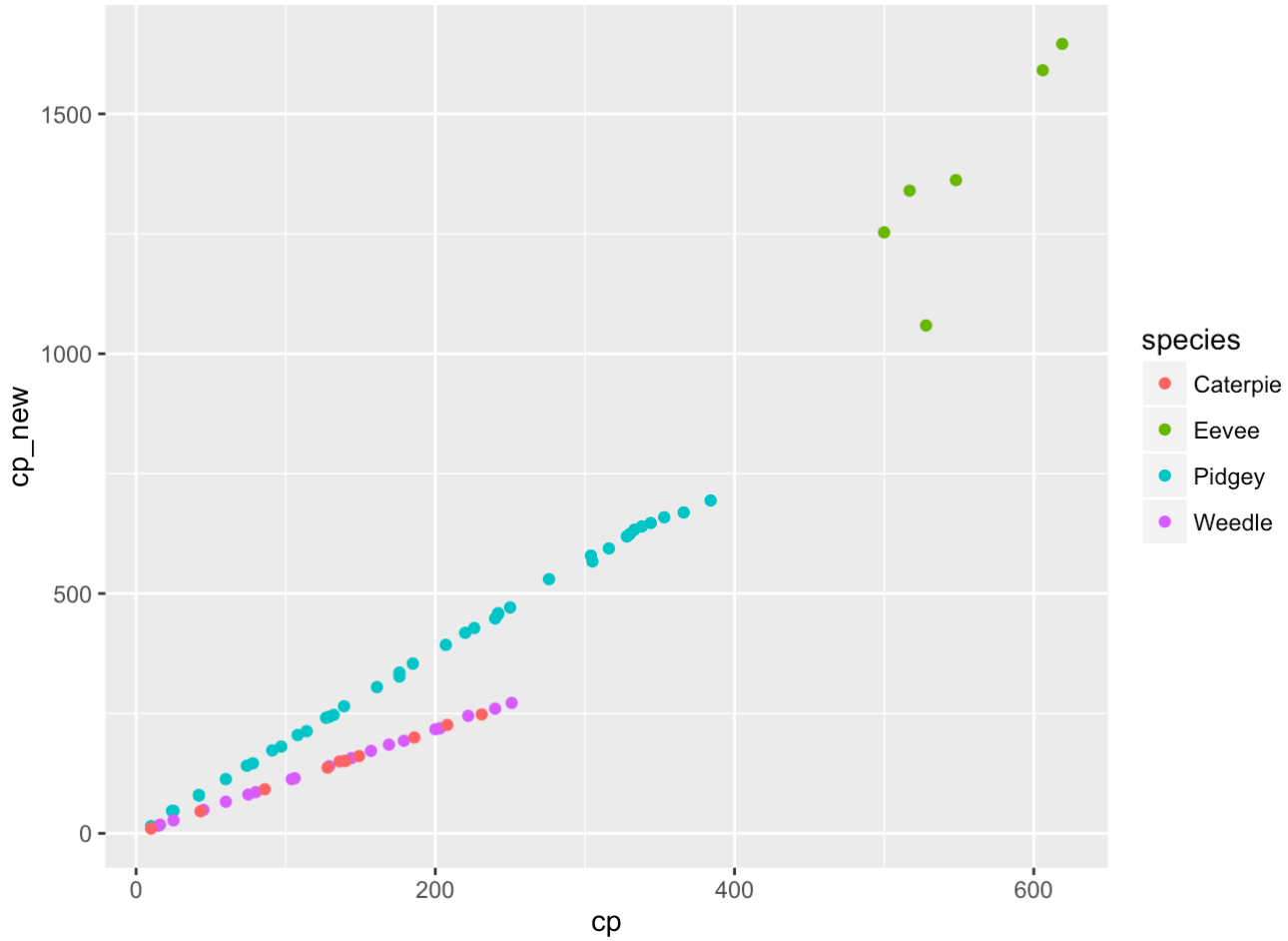
`ggplot2` plots are built up piece by piece. The following creates a blank plot, and store all the data to make that plot and build on it in the `myplot` object:

```
library(ggplot2)
myplot = ggplot(aes(cp, cp_new), data=pokemon)
```

A scatterplot

To add a scatterplot with colors and a legend, just add a `geom_point()` call. We could save this to a new object, like `myplot2` , but in this example we won’t create any new object, just make a scatterplot:

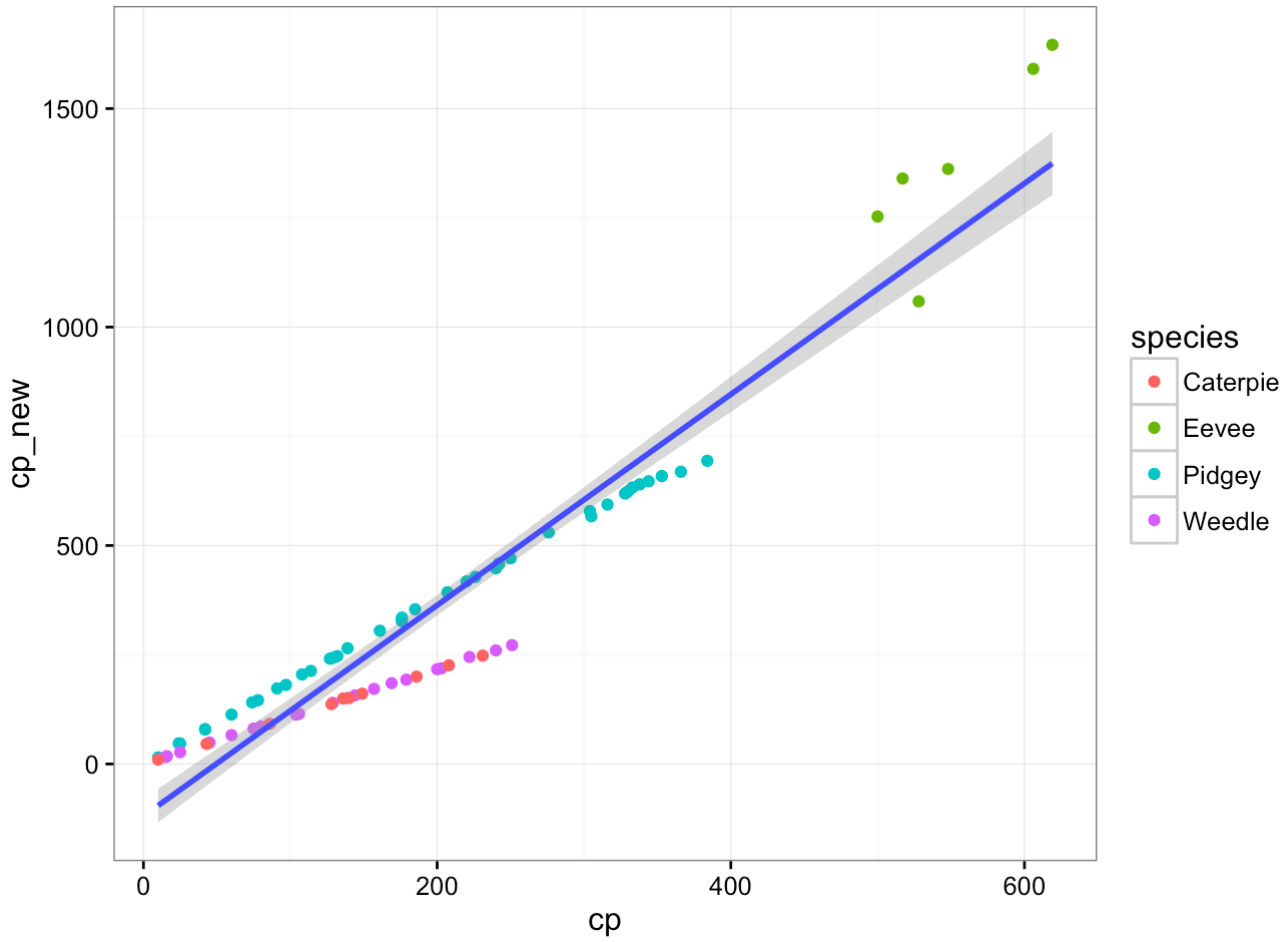
```
myplot + geom_point(aes(color=species))
```



A scatterplot plus regression lines

Try doing the following one by one, just adding new things to the existing plot. These functions are all documented in the `ggplot2` cheat sheet. You can also try skipping some of the lines. Note that nothing will happen until there is a line that doesn't end in "+"

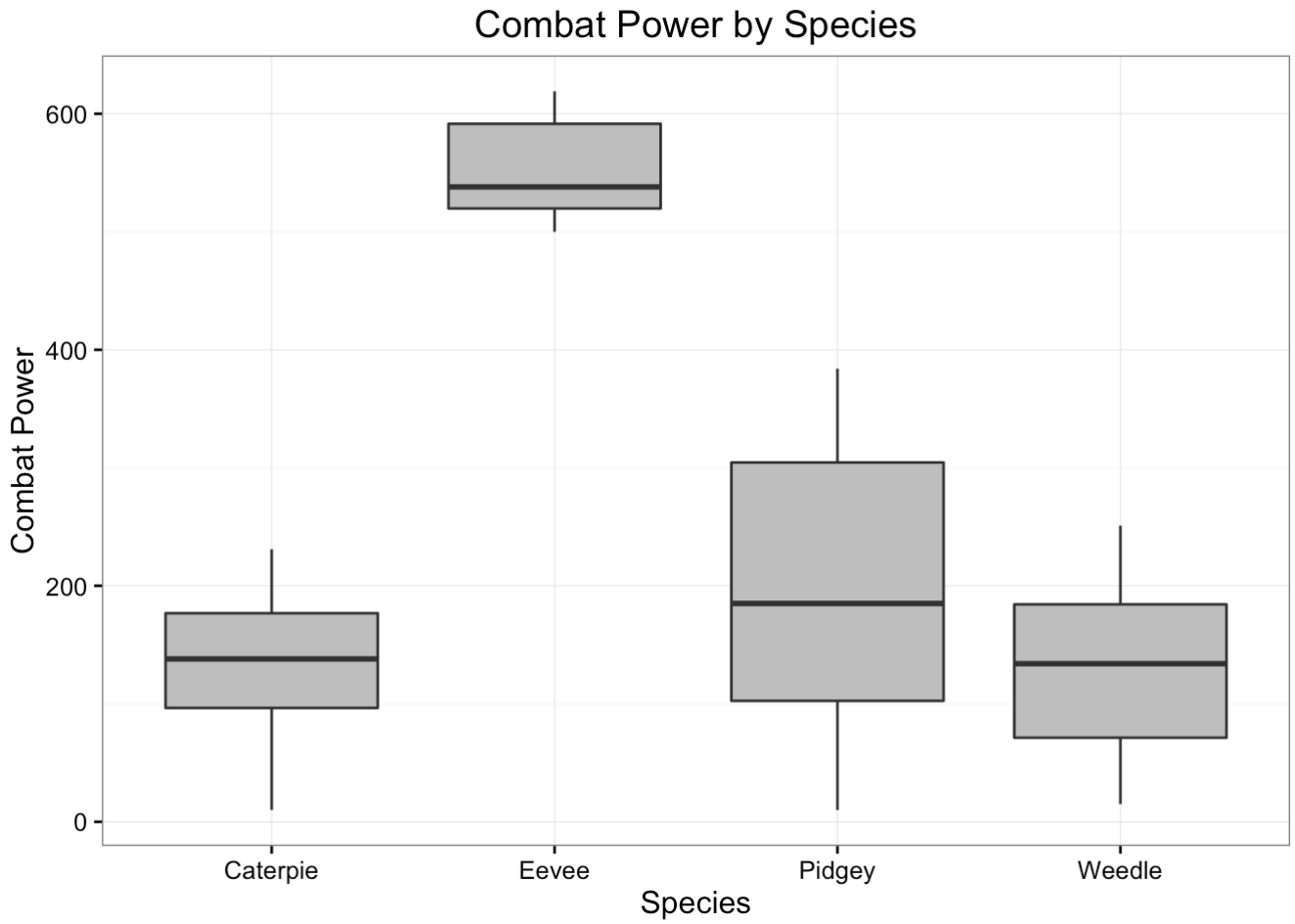
```
ggplot(aes(cp, cp_new), data=pokemon) +
  geom_point(aes(color=species)) +  #scatterplot
  geom_smooth(method="lm") +  #linear regression line and confidence bands
  theme_bw()  #get rid of the grey background
```



Boxplot

Now let's make a boxplot:

```
ggplot(aes(species, cp), data=pokemon) +
  geom_boxplot(fill="grey") +
  ggtitle("Combat Power by Species") +
  xlab("Species") +
  ylab("Combat Power") +
  theme_bw()
```



Try other kinds of plots, `geom_violin` and `geom_dotplot` ...