

Introduction to Tidyverse

Statlab Workshop

William Pang

Yale University

April 9, 2023

Pokemon!



Pokemon Dataset

Github Link

https://raw.githubusercontent.com/KeithGalli/pandas/master/pokemon_data.csv

Obtaining column specifications

```
# In your R code:  
spec(pokemon)
```

```
Rows: 800 Columns: 12
```

```
— Column specification —
```

```
Delimiter: ","
```

```
chr (3): Name, Type 1, Type 2
```

```
dbl (8): #, HP, Attack, Defense, Sp. Atk, Sp. Def, Speed, Generation
```

```
lgl (1): Legendary
```

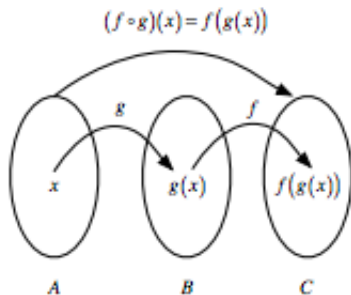
```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

The Pipe Operator

Key Idea

The pipe function takes the output of one function and passes it into another function as an argument



rowSums and the Pipe Operator

```
pokemon <- pokemon %>% mutate(total = rowSums(.[5:10]))  
pokemon <- dataframe -> mutate(rowSums(dataframe or vector?))
```

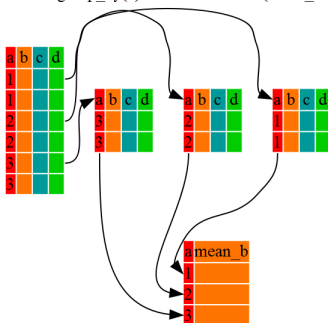


group_by

Key Idea

Collect data into groups and perform aggregate functions on the grouped data

```
data_frame %>% group_by(a) %>% summarize(mean_b=mean(b))
```



Wide and long data

Pokemon	HP	Attack	Defense
Bulbasaur	45	49	49
Venusaur	80	82	83
Chameleon	39	52	43

Pokemon	Ability	Value
Bulbasaur	HP	49
Bulbasaur	Attack	82
Bulbasaur	Defense	52
Venusaur	HP	80
Venusaur	Attack	82
Venusaur	Defense	83
Chameleon	HP	39
Chameleon	Attack	52
Chameleon	Defense	43

A soft introduction into ggplot

