RStudio for Data Science

David John Baker

06/04/2020

- Shiny
- Markdown
  - HTML
  - Word
  - LaTeX
- Terminal

Python

```
library(reticulate)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- tidyverse
```

```
## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  3.0.0     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------ tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
converted_tips_tidyverse <- read_csv("tips.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   total_bill = col_double(),
##   tip = col_double(),
##   sex = col_character(),
##   smoker = col_character(),
##   day = col_character(),
##   time = col_character(),
##   size = col_double()
## )
```

```
import pandas as pd
converted_tips_pandas = pd.read_csv("converted_tips.csv")
```

```
converted_tips_tidyverse
```

```
## # A tibble: 244 x 8
##       X1 total_bill   tip sex    smoker day   time    size
##    <dbl>      <dbl> <dbl> <chr>  <chr>  <chr> <chr>  <dbl>
##  1     0      17.0   1.01 Female No     Sun   Dinner     2
##  2     1      10.3   1.66 Male   No     Sun   Dinner     3
##  3     2      21.0   3.5  Male   No     Sun   Dinner     3
##  4     3      23.7   3.31 Male   No     Sun   Dinner     2
##  5     4      24.6   3.61 Female No     Sun   Dinner     4
##  6     5      25.3   4.71 Male   No     Sun   Dinner     4
##  7     6       8.77  2    Male   No     Sun   Dinner     2
##  8     7      26.9   3.12 Male   No     Sun   Dinner     4
##  9     8      15.0   1.96 Male   No     Sun   Dinner     2
## 10     9      14.8   3.23 Male   No     Sun   Dinner     2
## # ... with 234 more rows
```

```
converted_tips_tidyverse %>%
  select(day)
```

```
## # A tibble: 244 x 1
##    day
##    <chr>
##  1 Sun
##  2 Sun
##  3 Sun
##  4 Sun
##  5 Sun
##  6 Sun
##  7 Sun
##  8 Sun
##  9 Sun
## 10 Sun
## # ... with 234 more rows
```

```
converted_tips_pandas
```

```
##      total_bill   tip     sex smoker  ... yen_total yen_tips  aus_total  aus_tips
## 0         16.99  1.01  Female     No  ...   1800.94   107.06    10.8736    0.6464
## 1         10.34  1.66    Male     No  ...   1096.04   175.96     6.6176    1.0624
## 2         21.01  3.50    Male     No  ...   2227.06   371.00    13.4464    2.2400
## 3         23.68  3.31    Male     No  ...   2510.08   350.86    15.1552    2.1184
## 4         24.59  3.61  Female     No  ...   2606.54   382.66    15.7376    2.3104
## ..          ...   ...     ...    ...  ...       ...      ...        ...       ...
## 239       29.03  5.92    Male     No  ...   3077.18   627.52    18.5792    3.7888
## 240       27.18  2.00  Female    Yes  ...   2881.08   212.00    17.3952    1.2800
## 241       22.67  2.00    Male    Yes  ...   2403.02   212.00    14.5088    1.2800
## 242       17.82  1.75    Male     No  ...   1888.92   185.50    11.4048    1.1200
```

```
## 243       18.78  3.00  Female     No  ...   1990.68   318.00    12.0192    1.9200
##
## [244 rows x 13 columns]
```

```
converted_tips_pandas["day"]
```

```
## 0       Sun
## 1       Sun
## 2       Sun
## 3       Sun
## 4       Sun
##        ...
## 239     Sat
## 240     Sat
## 241     Sat
## 242     Sat
## 243    Thur
## Name: day, Length: 244, dtype: object
```

```
converted_tips_pandas["aus_sum"] = converted_tips_pandas["aus_tips"] + converted_tips_pandas["aus_total

converted_tips_pandas["aus_sum"]
```

```
## 0      11.5200
## 1       7.6800
## 2      15.6864
## 3      17.2736
## 4      18.0480
##          ...
## 239    22.3680
## 240    18.6752
## 241    15.7888
## 242    12.5248
## 243    13.9392
## Name: aus_sum, Length: 244, dtype: float64
```