Python Basics

R code

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(reticulate)
use_virtualenv("./.venv", required=TRUE)  #tells R reticulate to use this Python virtual envi
```

```
df_r <- read.csv("./data/covid.csv")
nrow(df_r)
```

```
[1] 20780
```

Python code

```
import pandas as pd

df_python_covid = pd.read_csv("./data/covid.csv")
df_python_covid.shape
```

```
(20780, 6)
```

```
df_python_covid.isna().sum()
```

```
date                   0
state               2153
tests                  0
cases                  0
hospitalizations       0
deaths                 0
dtype: int64
```

```
df_python_covid.dropna(inplace=True)
df_python_covid.isna().sum()
```

```
date                0
state               0
tests               0
cases               0
hospitalizations    0
deaths              0
dtype: int64
```

```
df_python_covid_agg = df_python_covid.groupby('state')['deaths'].sum()
df_python_covid_agg.index
```

```
Index(['AK', 'AL', 'AR', 'AZ', 'CA', 'CO', 'CT', 'DE', 'FL', 'GA', 'HI', 'IA',
       'ID', 'IL', 'IN', 'KS', 'KY', 'LA', 'MA', 'MD', 'ME', 'MI', 'MN', 'MO',
       'MS', 'MT', 'NC', 'ND', 'NE', 'NH', 'NJ', 'NM', 'NV', 'NY', 'OH', 'OK',
       'OR', 'PA', 'RI', 'SC', 'SD', 'TN', 'TX', 'UT', 'VA', 'VT', 'WA', 'WI',
       'WV', 'WY'],
      dtype='object', name='state')
```

How to exchange dataframes between Python and R

(Python to R)[https://stackoverflow.com/questions/74081417/how-to-visualize-a-pandas-dataframe-in-r-chunk-quarto] (R to Python)[https://www.r-bloggers.com/2023/01/combining-r-and-python-with-reticulate-and-quarto/]

```
nrow(reticulate::py$df_python_covid)
```

```
[1] 18627
```

```
back_to_r = reticulate::py$df_python_covid_agg
back_to_r
```

```
   AK    AL    AR    AZ    CA    CO    CT    DE    FL    GA    HI    IA    ID
  305 10148  5319 16328 54124  5989  7704  1473 32266 17906   445  5558  1879
   IL    IN    KS    KY    LA    MA    MD    ME    MI    MN    MO    MS    MT
23014 12737  4812  4819  9748 16417  7955   706 16658  6550  8161  6808  1381
   NC    ND    NE    NH    NJ    NM    NV    NY    OH    OK    OR    PA    RI
11502  1478  2113  1184 23574  3808  5037 39029 17656  4534  2296 24349  2547
   SC    SD    TN    TX    UT    VA    VT    WA    WI    WV    WY
 8754  1900 11543 44451  1976  9596   208  5041  7106  2325   682
```

```
type(r.df_r)
```

```
<class 'pandas.core.frame.DataFrame'>
```

Multi-index example

mpg - miles per gallon - is a dataset included in R

```
mpg
```

```
# A tibble: 234 x 11
   manufacturer model      displ  year   cyl trans drv     cty   hwy fl    class
   <chr>        <chr>      <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
 1 audi         a4           1.8  1999     4 auto~ f        18    29 p     comp~
 2 audi         a4           1.8  1999     4 manu~ f        21    29 p     comp~
 3 audi         a4           2    2008     4 manu~ f        20    31 p     comp~
 4 audi         a4           2    2008     4 auto~ f        21    30 p     comp~
 5 audi         a4           2.8  1999     6 auto~ f        16    26 p     comp~
 6 audi         a4           2.8  1999     6 manu~ f        18    26 p     comp~
 7 audi         a4           3.1  2008     6 auto~ f        18    27 p     comp~
 8 audi         a4 quattro   1.8  1999     4 manu~ 4        18    26 p     comp~
 9 audi         a4 quattro   1.8  1999     4 auto~ 4        16    25 p     comp~
10 audi         a4 quattro   2    2008     4 manu~ 4        20    28 p     comp~
# i 224 more rows
```

```
df_mpg = r.mpg
#df_mpg.shape
#df_mpg.info()
df_mpg_grouped = df_mpg.groupby(['manufacturer','class'])['year'].value_counts()
df_mpg_grouped_reset_index = df_mpg_grouped.reset_index()
print(df_mpg_grouped)
```

```
manufacturer  class       year
audi          compact     1999    8
                          2008    7
              midsize     2008    2
                          1999    1
chevrolet     2seater     2008    3
                                  ..
volkswagen    compact     2008    6
              midsize     1999    4
                          2008    3
              subcompact  1999    4
                          2008    2
Name: count, Length: 61, dtype: int64
```

```
print(df_mpg_grouped_reset_index)
```

```
   manufacturer       class  year  count
0          audi     compact  1999      8
1          audi     compact  2008      7
2          audi     midsize  2008      2
3          audi     midsize  1999      1
4     chevrolet     2seater  2008      3
..          ...         ...   ...    ...
56   volkswagen     compact  2008      6
57   volkswagen     midsize  1999      4
58   volkswagen     midsize  2008      3
59   volkswagen  subcompact  1999      4
60   volkswagen  subcompact  2008      2

[61 rows x 4 columns]
```

```
reticulate::py$df_mpg_grouped  #multi-index
```

```
      ('audi', 'compact', 1999)          ('audi', 'compact', 2008)
                              8                                  7
      ('audi', 'midsize', 2008)          ('audi', 'midsize', 1999)
                              2                                  1
 ('chevrolet', '2seater', 2008)     ('chevrolet', '2seater', 1999)
                              3                                  2
 ('chevrolet', 'midsize', 2008)     ('chevrolet', 'midsize', 1999)
                              3                                  2
     ('chevrolet', 'suv', 2008)         ('chevrolet', 'suv', 1999)
                              6                                  3
     ('dodge', 'minivan', 1999)         ('dodge', 'minivan', 2008)
                              6                                  5
      ('dodge', 'pickup', 2008)          ('dodge', 'pickup', 1999)
                             12                                  7
        ('dodge', 'suv', 2008)             ('dodge', 'suv', 1999)
                              4                                  3
       ('ford', 'pickup', 1999)           ('ford', 'pickup', 2008)
                              5                                  2
  ('ford', 'subcompact', 2008)       ('ford', 'subcompact', 1999)
                              5                                  4
         ('ford', 'suv', 1999)              ('ford', 'suv', 2008)
                              6                                  3
  ('honda', 'subcompact', 1999)      ('honda', 'subcompact', 2008)
                              5                                  4
   ('hyundai', 'midsize', 1999)       ('hyundai', 'midsize', 2008)
                              4                                  3
('hyundai', 'subcompact', 2008)    ('hyundai', 'subcompact', 1999)
                              5                                  2
          ('jeep', 'suv', 2008)              ('jeep', 'suv', 1999)
                              6                                  2
    ('land rover', 'suv', 1999)        ('land rover', 'suv', 2008)
                              2                                  2
       ('lincoln', 'suv', 1999)           ('lincoln', 'suv', 2008)
                              2                                  1
       ('mercury', 'suv', 1999)           ('mercury', 'suv', 2008)
                              2                                  2
    ('nissan', 'compact', 1999)        ('nissan', 'midsize', 2008)
                              2                                  5
    ('nissan', 'midsize', 1999)            ('nissan', 'suv', 1999)
                              2                                  2
        ('nissan', 'suv', 2008)        ('pontiac', 'midsize', 1999)
                              2                                  3
   ('pontiac', 'midsize', 2008)        ('subaru', 'compact', 2008)
```

```
                                2                                 4
    ('subaru', 'subcompact', 1999)            ('subaru', 'suv', 2008)
                                4                                 4
           ('subaru', 'suv', 1999)         ('toyota', 'compact', 1999)
                                2                                 7
       ('toyota', 'compact', 2008)         ('toyota', 'midsize', 1999)
                                5                                 4
       ('toyota', 'midsize', 2008)          ('toyota', 'pickup', 1999)
                                3                                 4
        ('toyota', 'pickup', 2008)             ('toyota', 'suv', 1999)
                                3                                 5
           ('toyota', 'suv', 2008)     ('volkswagen', 'compact', 1999)
                                3                                 8
   ('volkswagen', 'compact', 2008)     ('volkswagen', 'midsize', 1999)
                                6                                 4
   ('volkswagen', 'midsize', 2008) ('volkswagen', 'subcompact', 1999)
                                3                                 4
('volkswagen', 'subcompact', 2008)
                                2
```

```
print("---------")
```

```
[1] "---------"
```

```
reticulate::py$df_mpg_grouped_reset_index  #collapsed multi-index
```

```
   manufacturer      class year count
1          audi    compact 1999     8
2          audi    compact 2008     7
3          audi    midsize 2008     2
4          audi    midsize 1999     1
5     chevrolet    2seater 2008     3
6     chevrolet    2seater 1999     2
7     chevrolet    midsize 2008     3
8     chevrolet    midsize 1999     2
9     chevrolet        suv 2008     6
10    chevrolet        suv 1999     3
11        dodge    minivan 1999     6
12        dodge    minivan 2008     5
13        dodge     pickup 2008    12
14        dodge     pickup 1999     7
```

```
15        dodge        suv 2008     4
16        dodge        suv 1999     3
17         ford     pickup 1999     5
18         ford     pickup 2008     2
19         ford subcompact 2008     5
20         ford subcompact 1999     4
21         ford        suv 1999     6
22         ford        suv 2008     3
23        honda subcompact 1999     5
24        honda subcompact 2008     4
25      hyundai    midsize 1999     4
26      hyundai    midsize 2008     3
27      hyundai subcompact 2008     5
28      hyundai subcompact 1999     2
29         jeep        suv 2008     6
30         jeep        suv 1999     2
31   land rover        suv 1999     2
32   land rover        suv 2008     2
33      lincoln        suv 1999     2
34      lincoln        suv 2008     1
35      mercury        suv 1999     2
36      mercury        suv 2008     2
37       nissan    compact 1999     2
38       nissan    midsize 2008     5
39       nissan    midsize 1999     2
40       nissan        suv 1999     2
41       nissan        suv 2008     2
42      pontiac    midsize 1999     3
43      pontiac    midsize 2008     2
44       subaru    compact 2008     4
45       subaru subcompact 1999     4
46       subaru        suv 2008     4
47       subaru        suv 1999     2
48       toyota    compact 1999     7
49       toyota    compact 2008     5
50       toyota    midsize 1999     4
51       toyota    midsize 2008     3
52       toyota     pickup 1999     4
53       toyota     pickup 2008     3
54       toyota        suv 1999     5
55       toyota        suv 2008     3
56   volkswagen    compact 1999     8
57   volkswagen    compact 2008     6
```

```
58   volkswagen    midsize 1999     4
59   volkswagen    midsize 2008     3
60   volkswagen subcompact 1999     4
61   volkswagen subcompact 2008     2
```

```
#reticulate::py_last_error()
```