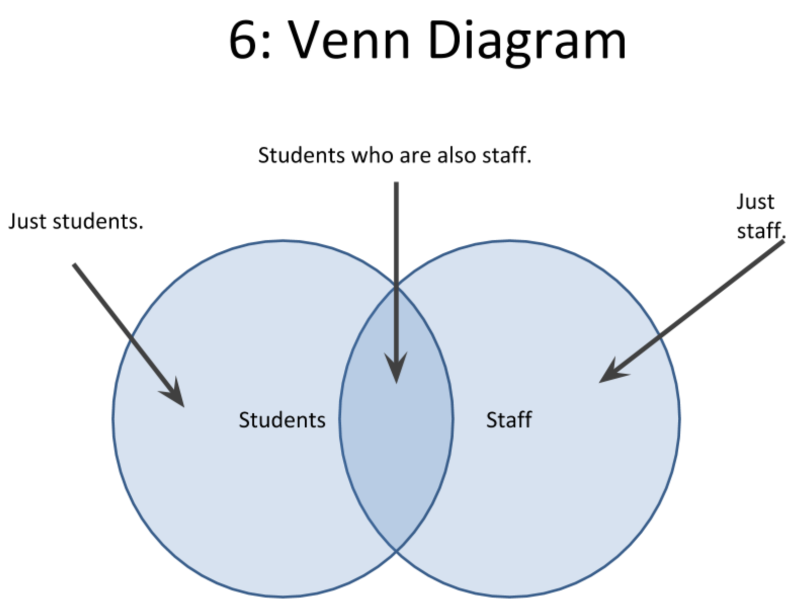
MergingDataFrame_ed

November 15, 2021

In this lecture we're going to address how you can bring multiple dataframe objects together, either by merging them horizontally, or by concatenating them vertically. Before we jump into the code, we need to address a little relational theory and to get some language conventions down. I'm going to bring in an image to help explain some concepts.



Venn Diagram

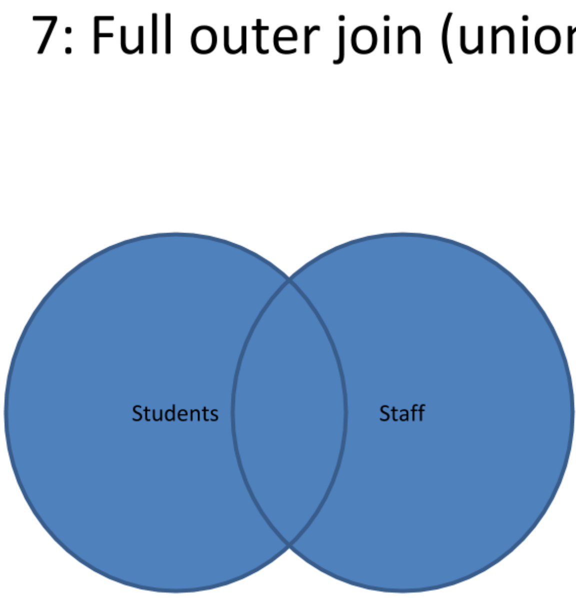
Ok, this is a Venn Diagram. A Venn Diagram is traditionally used to show set membership. For example, the circle on the left is the population of students at a university. The circle on the right is the population of staff at a university. And the overlapping region in the middle are all of those students who are also staff. Maybe these students run tutorials for a course, or grade assignments, or engage in running research experiments.

So, this diagram shows two populations whom we might have data about, but there is overlap between those populations.

When it comes to translating this to pandas, we can think of the case where we might have these two populations as indices in separate DataFrames, maybe with the label of Person Name.

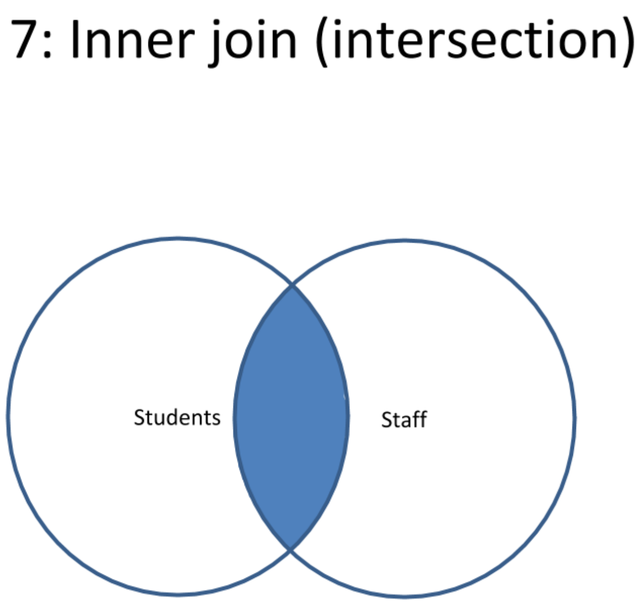
When we want to join the DataFrames together, we have some choices to make. First what if we want a list of all the people regardless of whether they're staff or student, and all of the information we can get on them? In database terminology, this is called a full outer join. And in set theory, it's called a union. In the Venn diagram, it represents everyone in any circle.

Here's an image of what that would look like in the Venn diagram.



Union

It's quite possible though that we only want those people who we have maximum information for, those people who are both staff and students. Maybe being a staff member and a student involves getting a tuition waiver, and we want to calculate the cost of this. In database terminology, this is called an inner join. Or in set theory, the intersection. It is represented in the Venn diagram as the overlapping parts of each circle.



Here's what that looks like:

[1]:
```python
# With that background, let's see an example of how we would do this in pandas, where we would use the merge
# function.
import pandas as pd

# First we create two DataFrames, staff and students.
staff_df = pd.DataFrame([{'Name': 'Kelly', 'Role': 'Director of HR'},
                         {'Name': 'Sally', 'Role': 'Course liasion'},
                         {'Name': 'James', 'Role': 'Grader'}])
# And lets index these staff by name
staff_df = staff_df.set_index('Name')
# Now we'll create a student dataframe
student_df = pd.DataFrame([{'Name': 'James', 'School': 'Business'},
                           {'Name': 'Mike', 'School': 'Law'},
                           {'Name': 'Sally', 'School': 'Engineering'}])
# And we'll index this by name too
student_df = student_df.set_index('Name')

# And lets just print out the dataframes
print(staff_df.head())
print(student_df.head())
```

```
                Role
Name
Kelly  Director of HR
Sally  Course liasion
James          Grader
```

```
             School
Name
James      Business
Mike            Law
Sally   Engineering
```

[2]:
```
# There's some overlap in these DataFrames in that James and Sally are both␣
↪students and staff, but Mike and
# Kelly are not. Importantly, both DataFrames are indexed along the value we␣
↪want to merge them on, which is
# called Name.
```

[3]:
```
# If we want the union of these, we would call merge() passing in the DataFrame␣
↪on the left and the DataFrame
# on the right and telling merge that we want it to use an outer join. We want␣
↪to use the left and right
# indices as the joining columns.

pd.merge(staff_df, student_df, how='outer', left_index=True, right_index=True)
```

[3]:
```
                   Role       School
Name
James            Grader     Business
Kelly    Director of HR          NaN
Mike                NaN          Law
Sally    Course liasion  Engineering
```

[4]:
```
# We see in the resulting DataFrame that everyone is listed. And since Mike␣
↪does not have a role, and John
# does not have a school, those cells are listed as missing values.

# If we wanted to get the intersection, that is, just those who are a student␣
↪AND a staff, we could set the
# how attribute to inner. Again, we set both left and right indices to be true␣
↪as the joining columns
pd.merge(staff_df, student_df, how='inner', left_index=True, right_index=True)
```

[4]:
```
                   Role       School
Name
Sally    Course liasion  Engineering
James            Grader     Business
```

[5]:
```
# And we see the resulting DataFrame has only James and Sally in it. Now there␣
↪are two other common use cases
# when merging DataFrames, and both are examples of what we would call set␣
↪addition. The first is when we
# would want to get a list of all staff regardless of whether they were␣
↪students or not. But if they were
```

```
# students, we would want to get their student details as well. To do this we
 ↪would use a left join. It is
# important to note the order of dataframes in this function: the first
 ↪dataframe is the left dataframe and
# the second is the right

pd.merge(staff_df, student_df, how='left', left_index=True, right_index=True)
```

[5]:

```
                 Role       School
Name
Kelly  Director of HR          NaN
Sally  Course liasion  Engineering
James          Grader     Business
```

[6]:

```
# You could probably guess what comes next. We want a list of all of the
 ↪students and their roles if they were
# also staff. To do this we would do a right join.
pd.merge(staff_df, student_df, how='right', left_index=True, right_index=True)
```

[6]:

```
                 Role       School
Name
James          Grader     Business
Mike              NaN          Law
Sally  Course liasion  Engineering
```

[7]:

```
# We can also do it another way. The merge method has a couple of other
 ↪interesting parameters. First, you
# don't need to use indices to join on, you can use columns as well. Here's an
 ↪example. Here we have a
# parameter called "on", and we can assign a column that both dataframe has as
 ↪the joining column

# First, lets remove our index from both of our dataframes
staff_df = staff_df.reset_index()
student_df = student_df.reset_index()

# Now lets merge using the on parameter
pd.merge(staff_df, student_df, how='right', on='Name')
```

[7]:

```
    Name            Role       School
0  Sally  Course liasion  Engineering
1  James          Grader     Business
2   Mike             NaN          Law
```

[8]:

```
# Using the "on" parameter instead of a the index is how I find myself using
 ↪merge() the most.
```

[9]:

```
# So what happens when we have conflicts between the DataFrames? Let's take a
 ↪look by creating new staff and
# student DataFrames that have a location information added to them.
```

```python
staff_df = pd.DataFrame([{'Name': 'Kelly', 'Role': 'Director of HR',
                          'Location': 'State Street'},
                         {'Name': 'Sally', 'Role': 'Course liasion',
                          'Location': 'Washington Avenue'},
                         {'Name': 'James', 'Role': 'Grader',
                          'Location': 'Washington Avenue'}])
student_df = pd.DataFrame([{'Name': 'James', 'School': 'Business',
                            'Location': '1024 Billiard Avenue'},
                           {'Name': 'Mike', 'School': 'Law',
                            'Location': 'Fraternity House #22'},
                           {'Name': 'Sally', 'School': 'Engineering',
                            'Location': '512 Wilson Crescent'}])

# In the staff DataFrame, this is an office location where we can find the␣
 ↪staff person. And we can see the
# Director of HR is on State Street, while the two students are on Washington␣
 ↪Avenue, and these locations just
# happen to be right outside my window as I film this. But for the student␣
 ↪DataFrame, the location information
# is actually their home address.

# The merge function preserves this information, but appends an _x or _y to␣
 ↪help differentiate between which
# index went with which column of data. The _x is always the left DataFrame␣
 ↪information, and the _y is always
# the right DataFrame information.

# Here, if we want all the staff information regardless of whether they were␣
 ↪students or not. But if they were
# students, we would want to get their student details as well.Then we can do a␣
 ↪left join and on the column of
# Name

pd.merge(staff_df, student_df, how='left', on='Name')
```

[9]:

	Name	Role	Location_x	School	Location_y
0	Kelly	Director of HR	State Street	NaN	NaN
1	Sally	Course liasion	Washington Avenue	Engineering	512 Wilson Crescent
2	James	Grader	Washington Avenue	Business	1024 Billiard Avenue

[10]:

```python
# From the output, we can see there are columns Location_x and Location_y.␣
 ↪Location_x refers to the Location
# column in the left dataframe, which is staff dataframe and Location_y refers␣
 ↪to the Location column in the
# right dataframe, which is student dataframe.
```

```
# Before we leave merging of DataFrames, let's talk about multi-indexing and␣
↪multiple columns. It's quite
# possible that the first name for students and staff might overlap, but the␣
↪last name might not. In this
# case, we use a list of the multiple columns that should be used to join keys␣
↪from both dataframes on the on
# parameter. Recall that the column name(s) assigned to the on parameter needs␣
↪to exist in both dataframes.

# Here's an example with some new student and staff data
staff_df = pd.DataFrame([{'First Name': 'Kelly', 'Last Name': 'Desjardins',
                          'Role': 'Director of HR'},
                         {'First Name': 'Sally', 'Last Name': 'Brooks',
                          'Role': 'Course liasion'},
                         {'First Name': 'James', 'Last Name': 'Wilde',
                          'Role': 'Grader'}])
student_df = pd.DataFrame([{'First Name': 'James', 'Last Name': 'Hammond',
                            'School': 'Business'},
                           {'First Name': 'Mike', 'Last Name': 'Smith',
                            'School': 'Law'},
                           {'First Name': 'Sally', 'Last Name': 'Brooks',
                            'School': 'Engineering'}])

# As you see here, James Wilde and James Hammond don't match on both keys since␣
↪they have different last
# names. So we would expect that an inner join doesn't include these␣
↪individuals in the output, and only Sally
# Brooks will be retained.
pd.merge(staff_df, student_df, how='inner', on=['First Name','Last Name'])
```

[10]:

```
  First Name Last Name            Role       School
0      Sally    Brooks  Course liasion  Engineering
```

[11]:

```
# Joining dataframes through merging is incredibly common, and you'll need to␣
↪know how to pull data from
# different sources, clean it, and join it for analysis. This is a staple not␣
↪only of pandas, but of database
# technologies as well.
```

[12]:

```
# If we think of merging as joining "horizontally", meaning we join on similar␣
↪values in a column found in two
# dataframes then concatenating is joining "vertically", meaning we put␣
↪dataframes on top or at the bottom of
# each other

# Let's understand this from an example. You have a dataset that tracks some␣
↪information over the years. And
```

```
# each year's record is a separate CSV and every CSV ofr every year's record␣
↪has the exactly same columns.
# What happens if you want to put all the data, from all years' record,␣
↪together? You can concatenate them.
```

[13]:
```
# Let's take a look at the US Department of Education College Scorecard data It␣
↪has each US university's data
# on student completion, student debt, after-graduation income, etc. The data␣
↪is stored in separate CSV's with
# each CSV containing a year's record Let's say we want the records from 2011␣
↪to 2013 we first create three
# dataframe, each containing one year's record. And, because the csv files␣
↪we're working with are messy, I
# want to supress some of the jupyter warning messages and just tell read_csv␣
↪to ignore bad lines, so I'm
# going to start the cell with a cell magic called %%capture
```

[14]:
```
%%capture
df_2011 = pd.read_csv("datasets/college_scorecard/MERGED2011_12_PP.csv",␣
↪error_bad_lines=False)
df_2012 = pd.read_csv("datasets/college_scorecard/MERGED2012_13_PP.csv",␣
↪error_bad_lines=False)
df_2013 = pd.read_csv("datasets/college_scorecard/MERGED2013_14_PP.csv",␣
↪error_bad_lines=False)
```

[15]:
```
# Let's get a view of one of the dataframes
df_2011.head(3)
```

[15]:
```
     UNITID      OPEID OPEID6                              INSTNM  \
0  100654.0   100200.0   1002            Alabama A & M University
1  100663.0   105200.0   1052  University of Alabama at Birmingham
2  100690.0  2503400.0  25034                  Amridge University

         CITY STABBR         ZIP  ACCREDAGENCY INSTURL NPCURL  ...  \
0      Normal     AL       35762           NaN     NaN    NaN  ...
1  Birmingham     AL  35294-0110           NaN     NaN    NaN  ...
2  Montgomery     AL  36117-3553           NaN     NaN    NaN  ...

  OMAWDP8_NOTFIRSTTIME_POOLED_SUPP OMENRUP_NOTFIRSTTIME_POOLED_SUPP  \
0                              NaN                              NaN
1                              NaN                              NaN
2                              NaN                              NaN

  OMENRYP_FULLTIME_POOLED_SUPP OMENRAP_FULLTIME_POOLED_SUPP  \
0                          NaN                          NaN
1                          NaN                          NaN
2                          NaN                          NaN
```

```
  OMAWDP8_FULLTIME_POOLED_SUPP  OMENRUP_FULLTIME_POOLED_SUPP  \
0                          NaN                           NaN
1                          NaN                           NaN
2                          NaN                           NaN

  OMENRYP_PARTTIME_POOLED_SUPP  OMENRAP_PARTTIME_POOLED_SUPP  \
0                          NaN                           NaN
1                          NaN                           NaN
2                          NaN                           NaN

  OMAWDP8_PARTTIME_POOLED_SUPP  OMENRUP_PARTTIME_POOLED_SUPP
0                          NaN                           NaN
1                          NaN                           NaN
2                          NaN                           NaN

[3 rows x 1977 columns]
```

[16]:

```
# We see that there is a whopping number of columns - more than 1900! We can␣
↪calculate the length of each
# dataframe as well
print(len(df_2011))
print(len(df_2012))
print(len(df_2013))
```

```
15235
7793
7804
```

[17]:

```
# That's a bit surprising that the number of schools in the scorecard for 2011␣
↪is almost double that of the
# next two years. But let's not worry about that. Instead, let's just put all␣
↪three dataframes in a list and
# call that list frames and pass the list into the concat() function Let's see␣
↪what it looks like

frames = [df_2011, df_2012, df_2013]
pd.concat(frames)
```

[17]:

```
          UNITID      OPEID  OPEID6  \
0       100654.0   100200.0    1002
1       100663.0   105200.0    1052
2       100690.0  2503400.0   25034
3       100706.0   105500.0    1055
4       100724.0   100500.0    1005
...          ...        ...     ...
7799  48285703.0   157107.0    1571
7800  48285704.0   157101.0    1571
7801  48285705.0   157105.0    1571
```

```
7802  48285706.0   157100.0   1571
7803  48285707.0   157103.0   1571

                                             INSTNM           CITY STABBR  \
0                          Alabama A & M University         Normal     AL
1               University of Alabama at Birmingham     Birmingham     AL
2                               Amridge University     Montgomery     AL
3               University of Alabama in Huntsville     Huntsville     AL
4                          Alabama State University     Montgomery     AL
...                                             ...            ...    ...
7799        Georgia Military College-Columbus Campus       Columbus     GA
7800        Georgia Military College-Valdosta Campus       Valdosta     GA
7801  Georgia Military College-Warner Robins Campus  Warner Robins     GA
7802              Georgia Military College-Online   Milledgeville     GA
7803         Georgia Military College-Stone Mountain  Stone Mountain     GA

             ZIP  ACCREDAGENCY INSTURL NPCURL  ...  \
0          35762           NaN     NaN    NaN  ...
1     35294-0110           NaN     NaN    NaN  ...
2     36117-3553           NaN     NaN    NaN  ...
3          35899           NaN     NaN    NaN  ...
4     36104-0271           NaN     NaN    NaN  ...
...          ...           ...     ...    ...  ...
7799       31909           NaN     NaN    NaN  ...
7800       31605           NaN     NaN    NaN  ...
7801       31093           NaN     NaN    NaN  ...
7802       31061           NaN     NaN    NaN  ...
7803       30083           NaN     NaN    NaN  ...

     OMAWDP8_NOTFIRSTTIME_POOLED_SUPP OMENRUP_NOTFIRSTTIME_POOLED_SUPP  \
0                                 NaN                              NaN
1                                 NaN                              NaN
2                                 NaN                              NaN
3                                 NaN                              NaN
4                                 NaN                              NaN
...                               ...                              ...
7799                              NaN                              NaN
7800                              NaN                              NaN
7801                              NaN                              NaN
7802                              NaN                              NaN
7803                              NaN                              NaN

     OMENRYP_FULLTIME_POOLED_SUPP OMENRAP_FULLTIME_POOLED_SUPP  \
0                             NaN                          NaN
1                             NaN                          NaN
2                             NaN                          NaN
3                             NaN                          NaN
```

```
4                               NaN                             NaN
...                             ...                             ...
7799                            NaN                             NaN
7800                            NaN                             NaN
7801                            NaN                             NaN
7802                            NaN                             NaN
7803                            NaN                             NaN

     OMAWDP8_FULLTIME_POOLED_SUPP OMENRUP_FULLTIME_POOLED_SUPP  \
0                               NaN                             NaN
1                               NaN                             NaN
2                               NaN                             NaN
3                               NaN                             NaN
4                               NaN                             NaN
...                             ...                             ...
7799                            NaN                             NaN
7800                            NaN                             NaN
7801                            NaN                             NaN
7802                            NaN                             NaN
7803                            NaN                             NaN

     OMENRYP_PARTTIME_POOLED_SUPP OMENRAP_PARTTIME_POOLED_SUPP  \
0                               NaN                             NaN
1                               NaN                             NaN
2                               NaN                             NaN
3                               NaN                             NaN
4                               NaN                             NaN
...                             ...                             ...
7799                            NaN                             NaN
7800                            NaN                             NaN
7801                            NaN                             NaN
7802                            NaN                             NaN
7803                            NaN                             NaN

     OMAWDP8_PARTTIME_POOLED_SUPP OMENRUP_PARTTIME_POOLED_SUPP
0                               NaN                             NaN
1                               NaN                             NaN
2                               NaN                             NaN
3                               NaN                             NaN
4                               NaN                             NaN
...                             ...                             ...
7799                            NaN                             NaN
7800                            NaN                             NaN
7801                            NaN                             NaN
7802                            NaN                             NaN
7803                            NaN                             NaN
```

```
[30832 rows x 1977 columns]
```

[18]:
```
# As you can see, we have more observations in one dataframe and columns remain␣
↪the same. If we scroll down to
# the bottom of the output, we see that there are a total of 30,832 rows after␣
↪concatenating three dataframes.
# Let's add the number of rows of the three dataframes and see if the two␣
↪numbers match
len(df_2011)+len(df_2012)+len(df_2013)
```

[18]:
```
30832
```

[19]:
```
# The two numbers match! Which means our concatenation is successful. But wait,␣
↪now that all the data is
# concatenated together, we don't know what observations are from what year␣
↪anymore! Actually the concat
# function has a parameter that solves such problem with the keys parameter, we␣
↪can set an extra level of
# indices, we pass in a list of keys that we want to correspond to the␣
↪dataframes into the keys parameter

# Now let's try it out
pd.concat(frames, keys=['2011','2012','2013'])
```

[19]:
```
                UNITID      OPEID  OPEID6  \
2011 0       100654.0   100200.0    1002
     1       100663.0   105200.0    1052
     2       100690.0  2503400.0   25034
     3       100706.0   105500.0    1055
     4       100724.0   100500.0    1005
...               ...        ...     ...
2013 7799  48285703.0   157107.0    1571
     7800  48285704.0   157101.0    1571
     7801  48285705.0   157105.0    1571
     7802  48285706.0   157100.0    1571
     7803  48285707.0   157103.0    1571

                                                INSTNM           CITY  \
2011 0                          Alabama A & M University         Normal
     1               University of Alabama at Birmingham     Birmingham
     2                               Amridge University     Montgomery
     3               University of Alabama in Huntsville     Huntsville
     4                          Alabama State University     Montgomery
...                                                  ...            ...
2013 7799       Georgia Military College-Columbus Campus       Columbus
     7800       Georgia Military College-Valdosta Campus       Valdosta
     7801  Georgia Military College-Warner Robins Campus  Warner Robins
     7802                Georgia Military College-Online   Milledgeville
```

```
     7803        Georgia Military College-Stone Mountain  Stone Mountain

          STABBR         ZIP  ACCREDAGENCY INSTURL NPCURL  ...  \
2011 0        AL       35762           NaN     NaN    NaN  ...
     1        AL  35294-0110           NaN     NaN    NaN  ...
     2        AL  36117-3553           NaN     NaN    NaN  ...
     3        AL       35899           NaN     NaN    NaN  ...
     4        AL  36104-0271           NaN     NaN    NaN  ...
...          ...         ...           ...     ...    ...  ...
2013 7799     GA       31909           NaN     NaN    NaN  ...
     7800     GA       31605           NaN     NaN    NaN  ...
     7801     GA       31093           NaN     NaN    NaN  ...
     7802     GA       31061           NaN     NaN    NaN  ...
     7803     GA       30083           NaN     NaN    NaN  ...

          OMAWDP8_NOTFIRSTTIME_POOLED_SUPP OMENRUP_NOTFIRSTTIME_POOLED_SUPP  \
2011 0                                 NaN                              NaN
     1                                 NaN                              NaN
     2                                 NaN                              NaN
     3                                 NaN                              NaN
     4                                 NaN                              NaN
...                                    ...                              ...
2013 7799                              NaN                              NaN
     7800                              NaN                              NaN
     7801                              NaN                              NaN
     7802                              NaN                              NaN
     7803                              NaN                              NaN

          OMENRYP_FULLTIME_POOLED_SUPP OMENRAP_FULLTIME_POOLED_SUPP  \
2011 0                             NaN                          NaN
     1                             NaN                          NaN
     2                             NaN                          NaN
     3                             NaN                          NaN
     4                             NaN                          NaN
...                                ...                          ...
2013 7799                          NaN                          NaN
     7800                          NaN                          NaN
     7801                          NaN                          NaN
     7802                          NaN                          NaN
     7803                          NaN                          NaN

          OMAWDP8_FULLTIME_POOLED_SUPP OMENRUP_FULLTIME_POOLED_SUPP  \
2011 0                             NaN                          NaN
     1                             NaN                          NaN
     2                             NaN                          NaN
     3                             NaN                          NaN
     4                             NaN                          NaN
```

```
...                                           ...                            ...
2013 7799                                     NaN                            NaN
     7800                                     NaN                            NaN
     7801                                     NaN                            NaN
     7802                                     NaN                            NaN
     7803                                     NaN                            NaN

          OMENRYP_PARTTIME_POOLED_SUPP OMENRAP_PARTTIME_POOLED_SUPP  \
2011 0                                NaN                            NaN
     1                                NaN                            NaN
     2                                NaN                            NaN
     3                                NaN                            NaN
     4                                NaN                            NaN
...                                   ...                            ...
2013 7799                             NaN                            NaN
     7800                             NaN                            NaN
     7801                             NaN                            NaN
     7802                             NaN                            NaN
     7803                             NaN                            NaN

          OMAWDP8_PARTTIME_POOLED_SUPP OMENRUP_PARTTIME_POOLED_SUPP
2011 0                                NaN                            NaN
     1                                NaN                            NaN
     2                                NaN                            NaN
     3                                NaN                            NaN
     4                                NaN                            NaN
...                                   ...                            ...
2013 7799                             NaN                            NaN
     7800                             NaN                            NaN
     7801                             NaN                            NaN
     7802                             NaN                            NaN
     7803                             NaN                            NaN

[30832 rows x 1977 columns]
```

[20]:
```
# Now we have the indices as the year so we know what observations are from␣
↪what year. You should know that
# concatenation also has inner and outer method. If you are concatenating two␣
↪dataframes that do not have
# identical columns, and choose the outer method, some cells will be NaN. If␣
↪you choose to do inner, then some
# observations will be dropped due to NaN values. You can think of this as␣
↪analogous to the left and right
# joins of the merge() function.
```

Now you know how to merge and concatenate datasets together. You will find such functions very useful for combining data to get more complex or complicated results and to do analysis with. A solid understanding of how to merge data is absolutely essentially when you are procuring, cleaning, and manipulating data. It's worth knowing how to join different datasets quickly, and the different options you can use when joining datasets, and I would encourage you to check out the pandas docs for joining and concatenating data.

PandasIdioms_ed

November 15, 2021

Python programmers will often suggest that there many ways the language can be used to solve a particular problem. But that some are more appropriate than others. The best solutions are celebrated as Idiomatic Python and there are lots of great examples of this on StackOverflow and other websites.

A sort of sub-language within Python, Pandas has its own set of idioms. We've alluded to some of these already, such as using vectorization whenever possible, and not using iterative loops if you don't need to. Several developers and users within the Panda's community have used the term **pandorable** for these idioms. I think it's a great term. So, I wanted to share with you a couple of key features of how you can make your code pandorable.

[1]:
```
# Let's start by bringing in our data processing libraries
import pandas as pd
import numpy as np
# And we'll bring in some timing functionality too, from the timeit module
import timeit

# And lets look at some census data from the US
df = pd.read_csv('datasets/census.csv')
df.head()
```

[1]:
```
   SUMLEV  REGION  DIVISION  STATE  COUNTY   STNAME          CTYNAME  \
0      40       3         6      1       0  Alabama          Alabama
1      50       3         6      1       1  Alabama   Autauga County
2      50       3         6      1       3  Alabama   Baldwin County
3      50       3         6      1       5  Alabama   Barbour County
4      50       3         6      1       7  Alabama      Bibb County

   CENSUS2010POP  ESTIMATESBASE2010  POPESTIMATE2010  ...  RDOMESTICMIG2011  \
0        4779736            4780127          4785161  ...          0.002295
1          54571              54571            54660  ...          7.242091
2         182265             182265           183193  ...         14.832960
3          27457              27457            27341  ...         -4.728132
4          22915              22919            22861  ...         -5.527043

   RDOMESTICMIG2012  RDOMESTICMIG2013  RDOMESTICMIG2014  RDOMESTICMIG2015  \
0         -0.193196          0.381066          0.582002         -0.467369
1         -2.915927         -3.012349          2.265971         -2.530799
2         17.647293         21.845705         19.243287         17.197872
```

```
3         -2.500690          -7.056824          -3.904217          -10.543299
4         -5.068871          -6.201001          -0.177537            0.177258

   RNETMIG2011  RNETMIG2012  RNETMIG2013  RNETMIG2014  RNETMIG2015
0     1.030015     0.826644     1.383282     1.724718     0.712594
1     7.606016    -2.626146    -2.722002     2.592270    -2.187333
2    15.844176    18.559627    22.727626    20.317142    18.293499
3    -4.874741    -2.758113    -7.167664    -3.978583   -10.543299
4    -5.088389    -4.363636    -5.403729     0.754533     1.107861

[5 rows x 100 columns]
```

[2]:

```
# The first of the pandas idioms I would like to talk about is called method␣
↪chaining. The general idea behind
# method chaining is that every method on an object returns a reference to that␣
↪object. The beauty of this is
# that you can condense many different operations on a DataFrame, for instance,␣
↪into one line or at least one
# statement of code.

# Here's the pandorable way to write code with method chaining. In this code␣
↪I'm going to pull out the state
# and city names as a multiple index, and I'm going to do so only for data␣
↪which has a summary level of 50,
# which in this dataset is county-level data. I'll rename a column too, just to␣
↪make it a bit more readable.
(df.where(df['SUMLEV']==50)
    .dropna()
    .set_index(['STNAME','CTYNAME'])
    .rename(columns={'ESTIMATESBASE2010': 'Estimates Base 2010'}))
```

[2]:

```
                          SUMLEV  REGION  DIVISION  STATE  COUNTY  \
STNAME  CTYNAME
Alabama Autauga County      50.0     3.0       6.0    1.0     1.0
        Baldwin County      50.0     3.0       6.0    1.0     3.0
        Barbour County      50.0     3.0       6.0    1.0     5.0
        Bibb County         50.0     3.0       6.0    1.0     7.0
        Blount County       50.0     3.0       6.0    1.0     9.0
...                          ...     ...       ...    ...     ...
Wyoming Sweetwater County   50.0     4.0       8.0   56.0    37.0
        Teton County        50.0     4.0       8.0   56.0    39.0
        Uinta County        50.0     4.0       8.0   56.0    41.0
        Washakie County     50.0     4.0       8.0   56.0    43.0
        Weston County       50.0     4.0       8.0   56.0    45.0

                          CENSUS2010POP  Estimates Base 2010  \
STNAME  CTYNAME
```

```
Alabama Autauga County          54571.0          54571.0
        Baldwin County         182265.0         182265.0
        Barbour County          27457.0          27457.0
        Bibb County             22915.0          22919.0
        Blount County           57322.0          57322.0
...                                 ...              ...
Wyoming Sweetwater County       43806.0          43806.0
        Teton County            21294.0          21294.0
        Uinta County            21118.0          21118.0
        Washakie County          8533.0           8533.0
        Weston County            7208.0           7208.0

                           POPESTIMATE2010  POPESTIMATE2011  POPESTIMATE2012  \
STNAME  CTYNAME
Alabama Autauga County             54660.0          55253.0          55175.0
        Baldwin County            183193.0         186659.0         190396.0
        Barbour County             27341.0          27226.0          27159.0
        Bibb County                22861.0          22733.0          22642.0
        Blount County              57373.0          57711.0          57776.0
...                                    ...              ...              ...
Wyoming Sweetwater County          43593.0          44041.0          45104.0
        Teton County               21297.0          21482.0          21697.0
        Uinta County               21102.0          20912.0          20989.0
        Washakie County             8545.0           8469.0           8443.0
        Weston County               7181.0           7114.0           7065.0

                           ...  RDOMESTICMIG2011  RDOMESTICMIG2012  \
STNAME  CTYNAME            ...
Alabama Autauga County     ...          7.242091         -2.915927
        Baldwin County     ...         14.832960         17.647293
        Barbour County     ...         -4.728132         -2.500690
        Bibb County        ...         -5.527043         -5.068871
        Blount County      ...          1.807375         -1.177622
...                        ...               ...               ...
Wyoming Sweetwater County  ...          1.072643         16.243199
        Teton County       ...         -1.589565          0.972695
        Uinta County       ...        -17.755986         -4.916350
        Washakie County    ...        -11.637475         -0.827815
        Weston County      ...        -11.752361         -8.040059

                           RDOMESTICMIG2013  RDOMESTICMIG2014  \
STNAME  CTYNAME
Alabama Autauga County            -3.012349          2.265971
        Baldwin County            21.845705         19.243287
        Barbour County            -7.056824         -3.904217
        Bibb County               -6.201001         -0.177537
        Blount County             -1.748766         -2.062535
```

```
...                                     ...             ...
Wyoming Sweetwater County         -5.339774      -14.252889
        Teton County              19.525929       14.143021
        Uinta County              -6.902954      -14.215862
        Washakie County           -2.013502      -17.781491
        Weston County             12.372583        1.533635

                           RDOMESTICMIG2015  RNETMIG2011  RNETMIG2012  \
STNAME  CTYNAME
Alabama Autauga County            -2.530799     7.606016    -2.626146
        Baldwin County            17.197872    15.844176    18.559627
        Barbour County           -10.543299    -4.874741    -2.758113
        Bibb County                0.177258    -5.088389    -4.363636
        Blount County             -1.369970     1.859511    -0.848580
...                                     ...          ...          ...
Wyoming Sweetwater County        -14.248864     1.255221    16.243199
        Teton County              -0.564849     0.654527     2.408578
        Uinta County             -12.127022   -18.136812    -5.536861
        Washakie County            1.682288   -11.990126    -1.182592
        Weston County              6.935294   -12.032179    -8.040059

                           RNETMIG2013  RNETMIG2014  RNETMIG2015
STNAME  CTYNAME
Alabama Autauga County       -2.722002     2.592270    -2.187333
        Baldwin County       22.727626    20.317142    18.293499
        Barbour County       -7.167664    -3.978583   -10.543299
        Bibb County          -5.403729     0.754533     1.107861
        Blount County        -1.402476    -1.577232    -0.884411
...                                ...          ...          ...
Wyoming Sweetwater County    -5.295460   -14.075283   -14.070195
        Teton County         21.160658    16.308671     1.520747
        Uinta County         -7.521840   -14.740608   -12.606351
        Washakie County      -2.250385   -18.020168     1.441961
        Weston County        12.372583     1.533635     6.935294

[3142 rows x 98 columns]
```

[3]:
```
# Lets walk through this. First, we use the where() function on the dataframe␣
↪and pass in a boolean mask which
# is only true for those rows where the SUMLEV is equal to 50. This indicates␣
↪in our source data that the data
# is summarized at the county level. With the result of the where() function␣
↪evaluated, we drop missing
# values. Remember that .where() doesn't drop missing values by default. Then␣
↪we set an index on the result of
# that. In this case I've set it to the state name followed by the county name.␣
↪Finally. I rename a column to
```

```
# make it more readable. Note that instead of writing this all on one line, as␣
↪I could have done, I began the
# statement with a parenthesis, which tells python I'm going to span the␣
↪statement over multiple lines for
# readability.
```

[4]:
```
# Here's a more traditional, non-pandorable way, of writing this. There's␣
↪nothing wrong with this code in the
# functional sense, you might even be able to understand it better as a new␣
↪person to the language. It's just
# not as pandorable as the first example.

# First create a new dataframe from the original
df = df[df['SUMLEV']==50] # I'll use the overloaded indexing operator [] which␣
↪drops nans
# Update the dataframe to have a new index, we use inplace=True to do this in␣
↪place
df.set_index(['STNAME','CTYNAME'], inplace=True)
# Set the column names
df.rename(columns={'ESTIMATESBASE2010': 'Estimates Base 2010'})
```

[4]:
```
                          SUMLEV  REGION  DIVISION  STATE  COUNTY  \
STNAME  CTYNAME
Alabama Autauga County        50       3         6      1       1
        Baldwin County        50       3         6      1       3
        Barbour County        50       3         6      1       5
        Bibb County           50       3         6      1       7
        Blount County         50       3         6      1       9
...                          ...     ...       ...    ...     ...
Wyoming Sweetwater County     50       4         8     56      37
        Teton County          50       4         8     56      39
        Uinta County          50       4         8     56      41
        Washakie County       50       4         8     56      43
        Weston County         50       4         8     56      45

                          CENSUS2010POP  Estimates Base 2010  \
STNAME  CTYNAME
Alabama Autauga County            54571                54571
        Baldwin County           182265               182265
        Barbour County            27457                27457
        Bibb County               22915                22919
        Blount County             57322                57322
...                                 ...                  ...
Wyoming Sweetwater County         43806                43806
        Teton County              21294                21294
        Uinta County              21118                21118
        Washakie County            8533                 8533
```

```
        Weston County              7208                7208

                                POPESTIMATE2010  POPESTIMATE2011  POPESTIMATE2012  \
STNAME  CTYNAME
Alabama Autauga County                    54660            55253            55175
        Baldwin County                   183193           186659           190396
        Barbour County                    27341            27226            27159
        Bibb County                       22861            22733            22642
        Blount County                     57373            57711            57776
...                                         ...              ...              ...
Wyoming Sweetwater County                 43593            44041            45104
        Teton County                      21297            21482            21697
        Uinta County                      21102            20912            20989
        Washakie County                    8545             8469             8443
        Weston County                      7181             7114             7065

                                ...  RDOMESTICMIG2011  RDOMESTICMIG2012  \
STNAME  CTYNAME                 ...
Alabama Autauga County          ...          7.242091         -2.915927
        Baldwin County          ...         14.832960         17.647293
        Barbour County          ...         -4.728132         -2.500690
        Bibb County             ...         -5.527043         -5.068871
        Blount County           ...          1.807375         -1.177622
...                             ...               ...               ...
Wyoming Sweetwater County       ...          1.072643         16.243199
        Teton County            ...         -1.589565          0.972695
        Uinta County            ...        -17.755986         -4.916350
        Washakie County         ...        -11.637475         -0.827815
        Weston County           ...        -11.752361         -8.040059

                                RDOMESTICMIG2013  RDOMESTICMIG2014  \
STNAME  CTYNAME
Alabama Autauga County                 -3.012349          2.265971
        Baldwin County                 21.845705         19.243287
        Barbour County                 -7.056824         -3.904217
        Bibb County                    -6.201001         -0.177537
        Blount County                  -1.748766         -2.062535
...                                          ...               ...
Wyoming Sweetwater County              -5.339774        -14.252889
        Teton County                   19.525929         14.143021
        Uinta County                   -6.902954        -14.215862
        Washakie County                -2.013502        -17.781491
        Weston County                  12.372583          1.533635

                                RDOMESTICMIG2015  RNETMIG2011  RNETMIG2012  \
STNAME  CTYNAME
Alabama Autauga County                 -2.530799     7.606016    -2.626146
```

```
        Baldwin County           17.197872   15.844176   18.559627
        Barbour County          -10.543299   -4.874741   -2.758113
        Bibb County               0.177258   -5.088389   -4.363636
        Blount County            -1.369970    1.859511   -0.848580
...                                    ...         ...         ...
Wyoming Sweetwater County       -14.248864    1.255221   16.243199
        Teton County             -0.564849    0.654527    2.408578
        Uinta County            -12.127022  -18.136812   -5.536861
        Washakie County           1.682288  -11.990126   -1.182592
        Weston County             6.935294  -12.032179   -8.040059

                           RNETMIG2013  RNETMIG2014  RNETMIG2015
STNAME  CTYNAME
Alabama Autauga County       -2.722002     2.592270    -2.187333
        Baldwin County       22.727626    20.317142    18.293499
        Barbour County       -7.167664    -3.978583   -10.543299
        Bibb County          -5.403729     0.754533     1.107861
        Blount County        -1.402476    -1.577232    -0.884411
...                                ...          ...          ...
Wyoming Sweetwater County    -5.295460   -14.075283   -14.070195
        Teton County         21.160658    16.308671     1.520747
        Uinta County         -7.521840   -14.740608   -12.606351
        Washakie County      -2.250385   -18.020168     1.441961
        Weston County        12.372583     1.533635     6.935294

[3142 rows x 98 columns]
```

[5]:
```
# Now, the key with any good idiom is to understand when it isn't helping you.␣
↪In this case, you can actually
# time both methods and see which one runs faster

# We can put the approach into a function and pass the function into the timeit␣
↪function to count the time the
# parameter number allows us to choose how many times we want to run the␣
↪function. Here we will just set it to
# 10

# Lets write a wrapper for our first function
def first_approach():
    global df
    # And we'll just paste our code right here
    return (df.where(df['SUMLEV']==50)
             .dropna()
             .set_index(['STNAME','CTYNAME'])
             .rename(columns={'ESTIMATESBASE2010': 'Estimates Base 2010'}))

# Read in our dataset anew
```

```
df = pd.read_csv('datasets/census.csv')

# And now lets run it
timeit.timeit(first_approach, number=10)
```

[5]: 1.1084833510685712

[6]:
```
# Now let's test the second approach. As you may notice, we use our global␣
↪variable df in the function.
# However, changing a global variable inside a function will modify the␣
↪variable even in a global scope and we
# do not want that to happen in this case. Therefore, for selecting summary␣
↪levels of 50 only, I create a new
# dataframe for those records

# Let's run this for once and see how fast it is
def second_approach():
    global df
    new_df = df[df['SUMLEV']==50]
    new_df.set_index(['STNAME','CTYNAME'], inplace=True)
    return new_df.rename(columns={'ESTIMATESBASE2010': 'Estimates Base 2010'})

# Read in our dataset anew
df = pd.read_csv('datasets/census.csv')

# And now lets run it
timeit.timeit(second_approach, number=10)
```

[6]: 0.10386669298168272

[7]:
```
# As you can see, the second approach is much faster! So, this is a particular␣
↪example of a classic time
# readability trade off.

# You'll see lots of examples on stack overflow and in documentation of people␣
↪using method chaining in their
# pandas. And so, I think being able to read and understand the syntax is␣
↪really worth your time. But keep in
# mind that following what appears to be stylistic idioms might have␣
↪performance issues that you need to
# consider as well.
```

[8]:
```
# Here's another pandas idiom. Python has a wonderful function called map,␣
↪which is sort of a basis for
# functional programming in the language. When you want to use map in Python,␣
↪you pass it some function you
# want called, and some iterable, like a list, that you want the function to be␣
↪applied to. The results are
```

```
# that the function is called against each item in the list, and there's a␣
↪resulting list of all of the
# evaluations of that function.

# Pandas has a similar function called applymap. In applymap, you provide some␣
↪function which should operate
# on each cell of a DataFrame, and the return set is itself a DataFrame. Now I␣
↪think applymap is fine, but I
# actually rarely use it. Instead, I find myself often wanting to map across␣
↪all of the rows in a DataFrame.
# And pandas has a function that I use heavily there, called apply. Let's look␣
↪at an example.
```

[9]:
```
# Let's take a look at our census DataFrame. In this DataFrame, we have five␣
↪columns for population estimates,
# with each column corresponding with one year of estimates. It's quite␣
↪reasonable to want to create some new
# columns for minimum or maximum values, and the apply function is an easy way␣
↪to do this.

# First, we need to write a function which takes in a particular row of data,␣
↪finds a minimum and maximum
# values, and returns a new row of data nd returns a new row of data.  We'll␣
↪call this function min_max, this
# is pretty straight forward. We can create some small slice of a row by␣
↪projecting the population columns.
# Then use the NumPy min and max functions, and create a new series with a␣
↪label values represent the new
# values we want to apply.

def min_max(row):
    data = row[['POPESTIMATE2010',
                'POPESTIMATE2011',
                'POPESTIMATE2012',
                'POPESTIMATE2013',
                'POPESTIMATE2014',
                'POPESTIMATE2015']]
    return pd.Series({'min': np.min(data), 'max': np.max(data)})
```

[10]:
```
# Then we just need to call apply on the DataFrame.

# Apply takes the function and the axis on which to operate as parameters. Now,␣
↪we have to be a bit careful,
# we've talked about axis zero being the rows of the DataFrame in the past. But␣
↪this parameter is really the
# parameter of the index to use. So, to apply across all rows, which is␣
↪applying on all columns, you pass axis
```

```
# equal to 'columns'.
df.apply(min_max, axis='columns').head()
```

[10]:

	min	max
0	4785161	4858979
1	54660	55347
2	183193	203709
3	26489	27341
4	22512	22861

[11]:
```
# Of course there's no need to limit yourself to returning a new series object.␣
↪If you're doing this as part
# of data cleaning your likely to find yourself wanting to add new data to the␣
↪existing DataFrame. In that
# case you just take the row values and add in new columns indicating the max␣
↪and minimum scores. This is a
# regular part of my workflow when bringing in data and building summary or␣
↪descriptive statistics, and is
# often used heavily with the merging of DataFrames.
```

[12]:
```
# Here's an example where we have a revised version of the function min_max␣
↪Instead of returning a separate
# series to display the min and max we add two new columns in the original␣
↪dataframe to store min and max

def min_max(row):
    data = row[['POPESTIMATE2010',
                'POPESTIMATE2011',
                'POPESTIMATE2012',
                'POPESTIMATE2013',
                'POPESTIMATE2014',
                'POPESTIMATE2015']]
    # Create a new entry for max
    row['max'] = np.max(data)
    # Create a new entry for min
    row['min'] = np.min(data)
    return row
# Now just apply the function across the dataframe
df.apply(min_max, axis='columns')
```

[12]:

	SUMLEV	REGION	DIVISION	STATE	COUNTY	STNAME	CTYNAME \
0	40	3	6	1	0	Alabama	Alabama
1	50	3	6	1	1	Alabama	Autauga County
2	50	3	6	1	3	Alabama	Baldwin County
3	50	3	6	1	5	Alabama	Barbour County
4	50	3	6	1	7	Alabama	Bibb County
...	...	...	...	...	...	...	...
3188	50	4	8	56	37	Wyoming	Sweetwater County

```
3189      50       4          8      56       39  Wyoming       Teton County
3190      50       4          8      56       41  Wyoming       Uinta County
3191      50       4          8      56       43  Wyoming    Washakie County
3192      50       4          8      56       45  Wyoming      Weston County

      CENSUS2010POP  ESTIMATESBASE2010  POPESTIMATE2010  ...  \
0           4779736            4780127          4785161  ...
1             54571              54571            54660  ...
2            182265             182265           183193  ...
3             27457              27457            27341  ...
4             22915              22919            22861  ...
...             ...                ...              ...  ...
3188          43806              43806            43593  ...
3189          21294              21294            21297  ...
3190          21118              21118            21102  ...
3191           8533               8533             8545  ...
3192           7208               7208             7181  ...

      RDOMESTICMIG2013  RDOMESTICMIG2014  RDOMESTICMIG2015  RNETMIG2011  \
0             0.381066          0.582002         -0.467369     1.030015
1            -3.012349          2.265971         -2.530799     7.606016
2            21.845705         19.243287         17.197872    15.844176
3            -7.056824         -3.904217        -10.543299    -4.874741
4            -6.201001         -0.177537          0.177258    -5.088389
...                ...               ...               ...          ...
3188         -5.339774        -14.252889        -14.248864     1.255221
3189         19.525929         14.143021         -0.564849     0.654527
3190         -6.902954        -14.215862        -12.127022   -18.136812
3191         -2.013502        -17.781491          1.682288   -11.990126
3192         12.372583          1.533635          6.935294   -12.032179

      RNETMIG2012  RNETMIG2013  RNETMIG2014  RNETMIG2015      max      min
0        0.826644     1.383282     1.724718     0.712594  4858979  4785161
1       -2.626146    -2.722002     2.592270    -2.187333    55347    54660
2       18.559627    22.727626    20.317142    18.293499   203709   183193
3       -2.758113    -7.167664    -3.978583   -10.543299    27341    26489
4       -4.363636    -5.403729     0.754533     1.107861    22861    22512
...           ...          ...          ...          ...      ...      ...
3188    16.243199    -5.295460   -14.075283   -14.070195    45162    43593
3189     2.408578    21.160658    16.308671     1.520747    23125    21297
3190    -5.536861    -7.521840   -14.740608   -12.606351    21102    20822
3191    -1.182592    -2.250385   -18.020168     1.441961     8545     8316
3192    -8.040059    12.372583     1.533635     6.935294     7234     7065

[3193 rows x 102 columns]
```

[13]:
```python
# Apply is an extremely important tool in your toolkit. The reason I introduced␣
↪apply here is because you
# rarely see it used with large function definitions, like we did. Instead, you␣
↪typically see it used with
# lambdas. To get the most of the discussions you'll see online, you're going␣
↪to need to know how to at least
# read lambdas.

# Here's You can imagine how you might chain several apply calls with lambdas␣
↪together to create a readable
# yet succinct data manipulation script. One line example of how you might␣
↪calculate the max of the columns
# using the apply function.
rows = ['POPESTIMATE2010', 'POPESTIMATE2011', 'POPESTIMATE2012',␣
↪'POPESTIMATE2013','POPESTIMATE2014',
        'POPESTIMATE2015']
# Now we'll just apply this across the dataframe with a lambda
df.apply(lambda x: np.max(x[rows]), axis=1).head()
```

[13]:
```
0    4858979
1      55347
2     203709
3      27341
4      22861
dtype: int64
```

[14]:
```python
# If you don't remember lambdas just pause the video for a moment and look up␣
↪the syntax. A lambda is just an
# unnamed function in python, in this case it takes a single parameter, x, and␣
↪returns a single value, in this
# case the maximum over all columns associated with row x.
```

[15]:
```python
# The beauty of the apply function is that it allows flexibility in doing␣
↪whatever manipulation that you
# desire, as the function you pass into apply can be any customized however you␣
↪want. Let's say we want to
# divide the states into four categories: Northeast, Midwest, South, and West␣
↪We can write a customized
# function that returns the region based on the state the state regions␣
↪information is obtained from Wikipedia

def get_state_region(x):
    northeast = ['Connecticut', 'Maine', 'Massachusetts', 'New Hampshire',
                 'Rhode Island','Vermont','New York','New␣
↪Jersey','Pennsylvania']
    midwest = ['Illinois','Indiana','Michigan','Ohio','Wisconsin','Iowa',
               'Kansas','Minnesota','Missouri','Nebraska','North Dakota',
```

```
                 'South Dakota']
    south = ['Delaware','Florida','Georgia','Maryland','North Carolina',
             'South Carolina','Virginia','District of Columbia','West Virginia',
             'Alabama','Kentucky','Mississippi','Tennessee','Arkansas',
             'Louisiana','Oklahoma','Texas']
    west = ['Arizona','Colorado','Idaho','Montana','Nevada','New Mexico','Utah',
            'Wyoming','Alaska','California','Hawaii','Oregon','Washington']

    if x in northeast:
        return "Northeast"
    elif x in midwest:
        return "Midwest"
    elif x in south:
        return "South"
    else:
        return "West"
```

[16]:
```
# Now we have the customized function, let's say we want to create a new column␣
↪called Region, which shows the
# state's region, we can use the customized function and the apply function to␣
↪do so. The customized function
# is supposed to work on the state name column STNAME. So we will set the apply␣
↪function on the state name
# column and pass the customized function into the apply function
df['state_region'] = df['STNAME'].apply(lambda x: get_state_region(x))
```

[17]:
```
# Now let's see the results
df[['STNAME','state_region']].head()
```

[17]:

	STNAME	state_region
0	Alabama	South
1	Alabama	South
2	Alabama	South
3	Alabama	South
4	Alabama	South

So there are a couple of Pandas idioms. But I think there's many more, and I haven't talked about them here. So here's an unofficial assignment for you. Go look at some of the top ranked questions on pandas on Stack Overflow, and look at how some of the more experienced authors, answer those questions. Do you see any interesting patterns? Feel free to share them with myself and others in the class.

GroupBy_ed

November 15, 2021

Sometimes we want to select data based on groups and understand aggregated data on a group level. We have seen that even though Pandas allows us to iterate over every row in a dataframe, it is geneally very slow to do so. Fortunately Pandas has a groupby() function to speed up such task. The idea behind the groupby() function is that it takes some dataframe, splits it into chunks based on some key values, applies computation on those chunks, then combines the results back together into another dataframe. In pandas this is refered to as the split-apply-combine pattern.

1 Splitting

[1]:
```
# Let's look at an example. First, we'll bring in our pandas and numpy␣
↪libraries
import pandas as pd
import numpy as np
```

[2]:
```
# Let's look at some US census data
df = pd.read_csv('datasets/census.csv')
# And exclude state level summarizations, which have sum level value of 40
df = df[df['SUMLEV']==50]
df.head()
```

[2]:
```
   SUMLEV  REGION  DIVISION  STATE  COUNTY   STNAME           CTYNAME  \
1      50       3         6      1       1  Alabama  Autauga County
2      50       3         6      1       3  Alabama  Baldwin County
3      50       3         6      1       5  Alabama  Barbour County
4      50       3         6      1       7  Alabama     Bibb County
5      50       3         6      1       9  Alabama   Blount County

   CENSUS2010POP  ESTIMATESBASE2010  POPESTIMATE2010  ...  RDOMESTICMIG2011  \
1          54571              54571            54660  ...          7.242091
2         182265             182265           183193  ...         14.832960
3          27457              27457            27341  ...         -4.728132
4          22915              22919            22861  ...         -5.527043
5          57322              57322            57373  ...          1.807375

   RDOMESTICMIG2012  RDOMESTICMIG2013  RDOMESTICMIG2014  RDOMESTICMIG2015  \
1         -2.915927         -3.012349          2.265971         -2.530799
2         17.647293         21.845705         19.243287         17.197872
```

```
3        -2.500690       -7.056824       -3.904217      -10.543299
4        -5.068871       -6.201001       -0.177537        0.177258
5        -1.177622       -1.748766       -2.062535       -1.369970

   RNETMIG2011  RNETMIG2012  RNETMIG2013  RNETMIG2014  RNETMIG2015
1     7.606016    -2.626146    -2.722002     2.592270    -2.187333
2    15.844176    18.559627    22.727626    20.317142    18.293499
3    -4.874741    -2.758113    -7.167664    -3.978583   -10.543299
4    -5.088389    -4.363636    -5.403729     0.754533     1.107861
5     1.859511    -0.848580    -1.402476    -1.577232    -0.884411

[5 rows x 100 columns]
```

[3]:
```
# In the first example for groupby() I want to use the census date. Let's get a␣
↪list of the unique states,
# then we can iterate over all the states and for each state we reduce the data␣
↪frame and calculate the
# average.

# Let's run such task for 3 times and time it. For this we'll use the cell␣
↪magic function %%timeit
```

[4]:
```
%%timeit -n 3

for state in df['STNAME'].unique():
    # We'll just calculate the average using numpy for this particular state
    avg = np.average(df.where(df['STNAME']==state).dropna()['CENSUS2010POP'])
    # And we'll print it to the screen
    print('Counties in state ' + state +
          ' have an average population of ' + str(avg))
```

```
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
```

```
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
```

```
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
```

```
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
```

```
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
```

```
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
```

```
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
```

```
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
```

```
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
```

```
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
```

```
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
```

```
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
```

```
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
```

```
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
```

```
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
```

```
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
```

```
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
```

```
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
```

```
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
```

```
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
```

```
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
```

```
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
```

```
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
```

```
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
4.62 s ś 230 ms per loop (mean ś std. dev. of 7 runs, 3 loops each)
```

[5]:
```
# If you scroll down to the bottom of that output you can see it takes a fair␣
↪bit of time to finish.
# Now let's try another approach using groupby()
```

[6]:
```
%%timeit -n 3
# For this method, we start by telling pandas we're interested in grouping by␣
↪state name, this is the "split"
for group, frame in df.groupby('STNAME'):
    # You'll notice there are two values we set here. groupby() returns a␣
↪tuple, where the first value is the
    # value of the key we were trying to group by, in this case a specific␣
↪state name, and the second one is
    # projected dataframe that was found for that group

    # Now we include our logic in the "apply" step, which is to calculate an␣
↪average of the census2010pop
    avg = np.average(frame['CENSUS2010POP'])
    # And print the results
    print('Counties in state ' + group +
          ' have an average population of ' + str(avg))
# And we don't have to worry about the combine step in this case, because all␣
↪of our data transformation is
# actually printing out results.
```

```
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
```

```
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
```

```
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
```

```
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
```

```
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
```

```
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
```

```
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
```

```
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
```

```
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
```

```
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
```

```
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
```

```
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
```

```
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
```

```
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
```

```
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
```

```
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
```

```
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
```

```
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
```

```
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
```

```
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
```

```
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
```

```
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
```

```
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
```

```
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of
100551.39130434782
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
29.6 ms ś 3.65 ms per loop (mean ś std. dev. of 7 runs, 3 loops each)
```

[7]:
```
# Wow, what a huge difference in speed. An improve by roughly by two factors!
```

[8]:
```
# Now, 99% of the time, you'll use group by on one or more columns. But you can also provide a function to
# group by and use that to segment your data.

# This is a bit of a fabricated example but lets say that you have a big batch job with lots of processing and
# you want to work on only a third or so of the states at a given time. We could create some function which
# returns a number between zero and two based on the first character of the state name. Then we can tell group
```

```
# by to use this function to split up our data frame. It's important to note␣
↪that in order to do this you need
# to set the index of the data frame to be the column that you want to group by␣
↪first.

# We'll create some new function called set_batch_number and if the first␣
↪letter of the parameter is a capital
# M we'll return a 0. If it's a capital Q we'll return a 1 and otherwise we'll␣
↪return a 2. Then we'll pass
# this function to the data frame

df = df.set_index('STNAME')

def set_batch_number(item):
    if item[0]<'M':
        return 0
    if item[0]<'Q':
        return 1
    return 2

# The dataframe is supposed to be grouped by according to the batch number And␣
↪we will loop through each batch
# group
for group, frame in df.groupby(set_batch_number):
    print('There are ' + str(len(frame)) + ' records in group ' + str(group) +␣
↪' for processing.')
```

```
There are 1177 records in group 0 for processing.
There are 1134 records in group 1 for processing.
There are 831 records in group 2 for processing.
```

[9]:
```
# Notice that this time I didn't pass in a column name to groupby(). Instead, I␣
↪set the index of the dataframe
# to be STNAME, and if no column identifier is passed groupby() will␣
↪automatically use the index.
```

[10]:
```
# Let's take one more look at an example of how we might group data. In this␣
↪example, I want to use a dataset
# of housing from airbnb. In this dataset there are two columns of interest,␣
↪one is the cancellation_policy
# and the other is the review_scores_value.
df=pd.read_csv("datasets/listings.csv")
df.head()
```

[10]:

	id	listing_url	scrape_id	\
0	12147973	https://www.airbnb.com/rooms/12147973	20160906204935	
1	3075044	https://www.airbnb.com/rooms/3075044	20160906204935	
2	6976	https://www.airbnb.com/rooms/6976	20160906204935	

```
3   1436513   https://www.airbnb.com/rooms/1436513  20160906204935
4   7651065   https://www.airbnb.com/rooms/7651065  20160906204935

  last_scraped                                          name  \
0   2016-09-07                    Sunny Bungalow in the City
1   2016-09-07             Charming room in pet friendly apt
2   2016-09-07              Mexican Folk Art Haven in Boston
3   2016-09-07  Spacious Sunny Bedroom Suite in Historic Home
4   2016-09-07                            Come Home to Boston

                                             summary  \
0  Cozy, sunny, family home.  Master bedroom high...
1  Charming and quiet room in a second floor 1910...
2  Come stay with a friendly, middle-aged guy in ...
3  Come experience the comforts of home away from...
4  My comfy, clean and relaxing home is one block...

                                               space  \
0  The house has an open and cozy feel at the sam...
1  Small but cozy and quite room with a full size...
2  Come stay with a friendly, middle-aged guy in ...
3  Most places you find in Boston are small howev...
4  Clean, attractive, private room, one block fro...

                                         description experiences_offered  \
0  Cozy, sunny, family home.  Master bedroom high...                none
1  Charming and quiet room in a second floor 1910...                none
2  Come stay with a friendly, middle-aged guy in ...                none
3  Come experience the comforts of home away from...                none
4  My comfy, clean and relaxing home is one block...                none

                               neighborhood_overview  ... review_scores_value  \
0  Roslindale is quiet, convenient and friendly. ...  ...                 NaN
1  The room is in Roslindale, a diverse and prima...  ...                 9.0
2  The LOCATION: Roslindale is a safe and diverse...  ...                10.0
3  Roslindale is a lovely little neighborhood loc...  ...                10.0
4  I love the proximity to downtown, the neighbor...  ...                10.0

  requires_license license jurisdiction_names instant_bookable  \
0                f     NaN                NaN                f
1                f     NaN                NaN                t
2                f     NaN                NaN                f
3                f     NaN                NaN                f
4                f     NaN                NaN                f

  cancellation_policy require_guest_profile_picture  \
0            moderate                             f
```

```
1          moderate                              f
2          moderate                              t
3          moderate                              f
4          flexible                              f

  require_guest_phone_verification calculated_host_listings_count  \
0                                f                              1
1                                f                              1
2                                f                              1
3                                f                              1
4                                f                              1

   reviews_per_month
0                NaN
1               1.30
2               0.47
3               1.00
4               2.25

[5 rows x 95 columns]
```

[11]:
```python
# So, how would I group by both of these columns? A first approach might be to␣
 ↪promote them to a multiindex
# and just call groupby()
df=df.set_index(["cancellation_policy","review_scores_value"])

# When we have a multiindex we need to pass in the levels we are interested in␣
 ↪grouping by
for group, frame in df.groupby(level=(0,1)):
    print(group)
```

```
('flexible', 2.0)
('flexible', 4.0)
('flexible', 5.0)
('flexible', 6.0)
('flexible', 7.0)
('flexible', 8.0)
('flexible', 9.0)
('flexible', 10.0)
('moderate', 2.0)
('moderate', 4.0)
('moderate', 6.0)
('moderate', 7.0)
('moderate', 8.0)
('moderate', 9.0)
('moderate', 10.0)
('strict', 2.0)
('strict', 3.0)
```

```
('strict', 4.0)
('strict', 5.0)
('strict', 6.0)
('strict', 7.0)
('strict', 8.0)
('strict', 9.0)
('strict', 10.0)
('super_strict_30', 6.0)
('super_strict_30', 7.0)
('super_strict_30', 8.0)
('super_strict_30', 9.0)
('super_strict_30', 10.0)
```

[12]:
```
# This seems to work ok. But what if we wanted to group by the cancelation␣
↪policy and review scores, but
# separate out all the 10's from those under ten? In this case, we could use a␣
↪function to manage the
# groupings

def grouping_fun(item):
    # Check the "review_scores_value" portion of the index. item is in the␣
↪format of
    # (cancellation_policy,review_scores_value
    if item[1] == 10.0:
        return (item[0],"10.0")
    else:
        return (item[0],"not 10.0")

for group, frame in df.groupby(by=grouping_fun):
    print(group)
```

```
('flexible', '10.0')
('flexible', 'not 10.0')
('moderate', '10.0')
('moderate', 'not 10.0')
('strict', '10.0')
('strict', 'not 10.0')
('super_strict_30', '10.0')
('super_strict_30', 'not 10.0')
```

[13]:
```
df.head()
```

[13]:
```
                                                id  \
cancellation_policy review_scores_value
moderate            NaN                   12147973
                    9.0                    3075044
                    10.0                      6976
                    10.0                   1436513
```

```
flexible            10.0                 7651065

                                                                     listing_url  
\
cancellation_policy review_scores_value                                           
moderate            NaN                 https://www.airbnb.com/rooms/12147973     
                    9.0                  https://www.airbnb.com/rooms/3075044     
                    10.0                    https://www.airbnb.com/rooms/6976     
                    10.0                 https://www.airbnb.com/rooms/1436513     
flexible            10.0                 https://www.airbnb.com/rooms/7651065     

                                             scrape_id last_scraped  \
cancellation_policy review_scores_value                               
moderate            NaN                 20160906204935   2016-09-07   
                    9.0                 20160906204935   2016-09-07   
                    10.0                20160906204935   2016-09-07   
                    10.0                20160906204935   2016-09-07   
flexible            10.0                20160906204935   2016-09-07   

  name  \
cancellation_policy review_scores_value                                           
moderate            NaN                                  Sunny Bungalow in 
the City   
                    9.0                           Charming room in pet 
friendly apt   
                    10.0                           Mexican Folk Art Haven in 
Boston   
                    10.0                Spacious Sunny Bedroom Suite in 
Historic Home   
flexible            10.0                                          Come Home to 
Boston   

   summary  \
cancellation_policy review_scores_value                                           
moderate            NaN                 Cozy, sunny, family home.  Master 
bedroom high...   
                    9.0                 Charming and quiet room in a second 
floor 1910...   
                    10.0                Come stay with a friendly, middle-aged 
guy in ...   
                    10.0                Come experience the comforts of home 
away from...   
flexible            10.0                My comfy, clean and relaxing home is 
one block...   

     space  \
cancellation_policy review_scores_value
```

```
moderate            NaN                     The house has an open and cozy feel at
the sam...
                    9.0                     Small but cozy and quite room with a
full size...
                    10.0                    Come stay with a friendly, middle-aged
guy in ...
                    10.0                    Most places you find in Boston are
small howev...
flexible            10.0                    Clean, attractive, private room, one
block fro...

description  \
cancellation_policy review_scores_value
moderate            NaN                     Cozy, sunny, family home.  Master
bedroom high...
                    9.0                     Charming and quiet room in a second
floor 1910...
                    10.0                    Come stay with a friendly, middle-aged
guy in ...
                    10.0                    Come experience the comforts of home
away from...
flexible            10.0                    My comfy, clean and relaxing home is
one block...

                                           experiences_offered  \
cancellation_policy review_scores_value
moderate            NaN                                   none
                    9.0                                   none
                    10.0                                  none
                    10.0                                  none
flexible            10.0                                  none

neighborhood_overview  \
cancellation_policy review_scores_value
moderate            NaN                     Roslindale is quiet, convenient and
friendly. ...
                    9.0                     The room is in Roslindale, a diverse
and prima...
                    10.0                    The LOCATION: Roslindale is a safe and
diverse...
                    10.0                    Roslindale is a lovely little
neighborhood loc...
flexible            10.0                    I love the proximity to downtown, the
neighbor...

                                            ... review_scores_communication  \
cancellation_policy review_scores_value  ...
```

```
moderate           NaN                    ...                          NaN
                   9.0                    ...                         10.0
                   10.0                   ...                         10.0
                   10.0                   ...                         10.0
flexible           10.0                   ...                         10.0

                                        review_scores_location  \
cancellation_policy review_scores_value
moderate           NaN                                     NaN
                   9.0                                     9.0
                   10.0                                    9.0
                   10.0                                   10.0
flexible           10.0                                    9.0

                                        requires_license license  \
cancellation_policy review_scores_value
moderate           NaN                                 f     NaN
                   9.0                                 f     NaN
                   10.0                                f     NaN
                   10.0                                f     NaN
flexible           10.0                                f     NaN

                                        jurisdiction_names instant_bookable  \
cancellation_policy review_scores_value
moderate           NaN                                 NaN                f
                   9.0                                 NaN                t
                   10.0                                NaN                f
                   10.0                                NaN                f
flexible           10.0                                NaN                f

                                        require_guest_profile_picture  \
cancellation_policy review_scores_value
moderate           NaN                                              f
                   9.0                                              f
                   10.0                                             t
                   10.0                                             f
flexible           10.0                                             f

                                        require_guest_phone_verification  \
cancellation_policy review_scores_value
moderate           NaN                                                 f
                   9.0                                                 f
                   10.0                                                f
                   10.0                                                f
flexible           10.0                                                f

                                        calculated_host_listings_count  \
```

```
cancellation_policy review_scores_value
moderate            NaN                                              1
                    9.0                                              1
                    10.0                                             1
                    10.0                                             1
flexible            10.0                                             1

                                        reviews_per_month
cancellation_policy review_scores_value
moderate            NaN                               NaN
                    9.0                              1.30
                    10.0                             0.47
                    10.0                             1.00
flexible            10.0                             2.25

[5 rows x 93 columns]
```

2 Applying

[14]:
```
# To this point we have applied very simple processing to our data after␣
↪splitting, really just outputting
# some print statements to demonstrate how the splitting works. The pandas␣
↪developers have three broad
# categories of data processing to happen during the apply step, Aggregation of␣
↪group data, Transformation of
# group data, and Filtration of group data
```

2.1 Aggregation

[15]:
```
# The most straight forward apply step is the aggregation of data, and uses the␣
↪method agg() on the groupby
# object. Thus far we have only iterated through the groupby object, unpacking␣
↪it into a label (the group
# name) and a dataframe. But with agg we can pass in a dictionary of the␣
↪columns we are interested in
# aggregating along with the function we are looking to apply to aggregate.

# Let's reset the index for our airbnb data
df=df.reset_index()

# Now lets group by the cancellation policy and find the average␣
↪review_scores_value by group
df.groupby("cancellation_policy").agg({"review_scores_value":np.average})
```

[15]:
```
                     review_scores_value
cancellation_policy
flexible                             NaN
moderate                             NaN
strict                               NaN
super_strict_30                      NaN
```

[16]:
```
# Hrm. That didn't seem to work at all. Just a bunch of not a numbers. The␣
↪issue is actually in the function
# that we sent to aggregate. np.average does not ignore nans! However, there is␣
↪a function we can use for this
df.groupby("cancellation_policy").agg({"review_scores_value":np.nanmean})
```

[16]:
```
                     review_scores_value
cancellation_policy
flexible                        9.237421
moderate                        9.307398
strict                          9.081441
super_strict_30                 8.537313
```

[17]:
```
# We can just extend this dictionary to aggregate by multiple functions or␣
↪multiple columns.
df.groupby("cancellation_policy").agg({"review_scores_value":(np.nanmean,np.
↪nanstd),
                                      "reviews_per_month":np.nanmean})
```

[17]:
```
                    review_scores_value           reviews_per_month
                                nanmean    nanstd           nanmean
cancellation_policy
flexible                       9.237421  1.096271          1.829210
moderate                       9.307398  0.859859          2.391922
strict                         9.081441  1.040531          1.873467
super_strict_30                8.537313  0.840785          0.340143
```

[18]:
```
# Take a moment to make sure you understand the previous cell, since it's␣
↪somewhat complex. First we're doing
# a group by on the dataframe object by the column "cancellation_policy". This␣
↪creates a new GroupBy object.
# Then we are invoking the agg() function on that object. The agg function is␣
↪going to apply one or more
# functions we specify to the group dataframes and return a single row per␣
↪dataframe/group. When we called
# this function we sent it two dictionary entries, each with the key indicating␣
↪which column we wanted
# functions applied to. For the first column we actually supplied a tuple of␣
↪two functions. Note that these
# are not function invocations, like np.nanmean(), or function names, like␣
↪"nanmean" they are references to
```

```
# functions which will return single values. The groupby object will recognize␣
↪the tuple and call each
# function in order on the same column. The results will be in a heirarchical␣
↪index, but since they are
# columns they don't show as an index per se. Then we indicated another column␣
↪and a single function we wanted
# to run.
```

2.2 Transformation

[19]:
```
# Transformation is different from aggregation. Where agg() returns a single␣
↪value per column, so one row per
# group, tranform() returns an object that is the same size as the group.␣
↪Essentially, it broadcasts the
# function you supply over the grouped dataframe, returning a new dataframe.␣
↪This makes combining data later
# easy.
```

[20]:
```
# For instance, suppose we want to include the average rating values in a given␣
↪group by cancellation policy,
# but preserve the dataframe shape so that we could generate a difference␣
↪between an individual observation
# and the sum.

# First, lets define just some subset of columns we are interested in
cols=['cancellation_policy','review_scores_value']
# Now lets transform it, I'll store this in its own dataframe
transform_df=df[cols].groupby('cancellation_policy').transform(np.nanmean)
transform_df.head()
```

[20]:
```
   review_scores_value
0             9.307398
1             9.307398
2             9.307398
3             9.307398
4             9.237421
```

[21]:
```
# So we can see that the index here is actually the same as the original␣
↪dataframe. So lets just join this
# in. Before we do that, lets rename the column in the transformed version
transform_df.rename({'review_scores_value':'mean_review_scores'},␣
↪axis='columns', inplace=True)
df=df.merge(transform_df, left_index=True, right_index=True)
df.head()
```

[21]:
```
  cancellation_policy  review_scores_value        id  \
0            moderate                  NaN  12147973
1            moderate                  9.0   3075044
```

```
2           moderate                 10.0      6976
3           moderate                 10.0   1436513
4           flexible                 10.0   7651065

                           listing_url       scrape_id last_scraped  \
0  https://www.airbnb.com/rooms/12147973  20160906204935   2016-09-07
1   https://www.airbnb.com/rooms/3075044  20160906204935   2016-09-07
2      https://www.airbnb.com/rooms/6976  20160906204935   2016-09-07
3   https://www.airbnb.com/rooms/1436513  20160906204935   2016-09-07
4   https://www.airbnb.com/rooms/7651065  20160906204935   2016-09-07

                                          name  \
0                   Sunny Bungalow in the City
1            Charming room in pet friendly apt
2             Mexican Folk Art Haven in Boston
3  Spacious Sunny Bedroom Suite in Historic Home
4                          Come Home to Boston

                                             summary  \
0  Cozy, sunny, family home.  Master bedroom high...
1  Charming and quiet room in a second floor 1910...
2  Come stay with a friendly, middle-aged guy in ...
3  Come experience the comforts of home away from...
4  My comfy, clean and relaxing home is one block...

                                               space  \
0  The house has an open and cozy feel at the sam...
1  Small but cozy and quite room with a full size...
2  Come stay with a friendly, middle-aged guy in ...
3  Most places you find in Boston are small howev...
4  Clean, attractive, private room, one block fro...

                                         description  ...  \
0  Cozy, sunny, family home.  Master bedroom high...  ...
1  Charming and quiet room in a second floor 1910...  ...
2  Come stay with a friendly, middle-aged guy in ...  ...
3  Come experience the comforts of home away from...  ...
4  My comfy, clean and relaxing home is one block...  ...

  review_scores_location requires_license license jurisdiction_names  \
0                    NaN                f     NaN                NaN
1                    9.0                f     NaN                NaN
2                    9.0                f     NaN                NaN
3                   10.0                f     NaN                NaN
4                    9.0                f     NaN                NaN

  instant_bookable require_guest_profile_picture  \
```

```
0               f                            f
1               t                            f
2               f                            t
3               f                            f
4               f                            f

  require_guest_phone_verification calculated_host_listings_count  \
0                                f                              1
1                                f                              1
2                                f                              1
3                                f                              1
4                                f                              1

  reviews_per_month mean_review_scores
0               NaN           9.307398
1              1.30           9.307398
2              0.47           9.307398
3              1.00           9.307398
4              2.25           9.237421

[5 rows x 96 columns]
```

[22]:
```
# Great, we can see that our new column is in place, the mean_review_scores. So␣
↪now we could create, for
# instance, the difference between a given row and it's group (the cancellation␣
↪policy) means.
df['mean_diff']=np.absolute(df['review_scores_value']-df['mean_review_scores'])
df['mean_diff'].head()
```

[22]:
```
0         NaN
1    0.307398
2    0.692602
3    0.692602
4    0.762579
Name: mean_diff, dtype: float64
```

2.3 Filtering

[23]:
```
# The GroupBy object has build in support for filtering groups as well. It's␣
↪often that you'll want to group
# by some feature, then make some transformation to the groups, then drop␣
↪certain groups as part of your
# cleaning routines. The filter() function takes in a function which it applies␣
↪to each group dataframe and
# returns either a True or a False, depending upon whether that group should be␣
↪included in the results.
```

[24]:
```
# For instance, if we only want those groups which have a mean rating above 9␣
↪included in our results
df.groupby('cancellation_policy').filter(lambda x: np.
↪nanmean(x['review_scores_value'])>9.2)
```

[24]:
```
     cancellation_policy  review_scores_value        id  \
0               moderate                  NaN  12147973
1               moderate                  9.0   3075044
2               moderate                 10.0      6976
3               moderate                 10.0   1436513
4               flexible                 10.0   7651065
...                  ...                  ...       ...
3576            flexible                  NaN  14689681
3577            flexible                  NaN  13750763
3579            flexible                  NaN  14852179
3582            flexible                  NaN  14585486
3584            flexible                  NaN  14504422

                                   listing_url       scrape_id last_scraped  \
0     https://www.airbnb.com/rooms/12147973  20160906204935   2016-09-07
1      https://www.airbnb.com/rooms/3075044  20160906204935   2016-09-07
2         https://www.airbnb.com/rooms/6976  20160906204935   2016-09-07
3      https://www.airbnb.com/rooms/1436513  20160906204935   2016-09-07
4      https://www.airbnb.com/rooms/7651065  20160906204935   2016-09-07
...                                    ...             ...          ...
3576  https://www.airbnb.com/rooms/14689681  20160906204935   2016-09-07
3577  https://www.airbnb.com/rooms/13750763  20160906204935   2016-09-07
3579  https://www.airbnb.com/rooms/14852179  20160906204935   2016-09-07
3582  https://www.airbnb.com/rooms/14585486  20160906204935   2016-09-07
3584  https://www.airbnb.com/rooms/14504422  20160906204935   2016-09-07

                                                  name  \
0                           Sunny Bungalow in the City
1                    Charming room in pet friendly apt
2                     Mexican Folk Art Haven in Boston
3          Spacious Sunny Bedroom Suite in Historic Home
4                                  Come Home to Boston
...                                                ...
3576      Beautiful loft style bedroom with large bathroom
3577           Comfortable Space in the Heart of Brookline
3579  Spacious Queen Bed Room Close to Boston Univer...
3582                           Gorgeous funky apartment
3584                    (K1) Private Room near Harvard/MIT

                                               summary  \
0     Cozy, sunny, family home.  Master bedroom high...
1     Charming and quiet room in a second floor 1910...
```

```
2     Come stay with a friendly, middle-aged guy in ...
3     Come experience the comforts of home away from...
4     My comfy, clean and relaxing home is one block...
...                                                ...
3576  You'd be living on the top floor of a four sto...
3577  Our place is close to Coolidge Corner, Allston...
3579  - Grocery: A full-size Star market is 2 minute...
3582  Funky little apartment close to public transpo...
3584  My place is close to My home is a warm and fri...

                                                 space  \
0     The house has an open and cozy feel at the sam...
1     Small but cozy and quite room with a full size...
2     Come stay with a friendly, middle-aged guy in ...
3     Most places you find in Boston are small howev...
4     Clean, attractive, private room, one block fro...
...                                                ...
3576                                               NaN
3577  This space consists of 2 Rooms and a private b...
3579                                               NaN
3582  Modern and relaxed space with many facilities ...
3584  To ensure a smooth check in: 1. You MUST have ...

                                           description  ... requires_license  \
0     Cozy, sunny, family home.  Master bedroom high...  ...                f
1     Charming and quiet room in a second floor 1910...  ...                f
2     Come stay with a friendly, middle-aged guy in ...  ...                f
3     Come experience the comforts of home away from...  ...                f
4     My comfy, clean and relaxing home is one block...  ...                f
...                                                ...  ...              ...
3576  You'd be living on the top floor of a four sto...  ...                f
3577  Our place is close to Coolidge Corner, Allston...  ...                f
3579  - Grocery: A full-size Star market is 2 minute...  ...                f
3582  Funky little apartment close to public transpo...  ...                f
3584  My place is close to My home is a warm and fri...  ...                f

     license jurisdiction_names instant_bookable  \
0        NaN                NaN                f
1        NaN                NaN                t
2        NaN                NaN                f
3        NaN                NaN                f
4        NaN                NaN                f
...      ...                ...              ...
3576     NaN                NaN                f
3577     NaN                NaN                f
3579     NaN                NaN                f
3582     NaN                NaN                f
```

```
3584     NaN                 NaN                 t

     require_guest_profile_picture require_guest_phone_verification  \
0                                f                                f
1                                f                                f
2                                t                                f
3                                f                                f
4                                f                                f
...                            ...                              ...
3576                             f                                f
3577                             f                                f
3579                             f                                f
3582                             f                                f
3584                             f                                f

     calculated_host_listings_count reviews_per_month mean_review_scores  \
0                                 1               NaN           9.307398
1                                 1              1.30           9.307398
2                                 1              0.47           9.307398
3                                 1              1.00           9.307398
4                                 1              2.25           9.237421
...                             ...               ...                ...
3576                              1               NaN           9.237421
3577                              1               NaN           9.237421
3579                              1               NaN           9.237421
3582                              1               NaN           9.237421
3584                              3               NaN           9.237421

     mean_diff
0          NaN
1     0.307398
2     0.692602
3     0.692602
4     0.762579
...        ...
3576       NaN
3577       NaN
3579       NaN
3582       NaN
3584       NaN

[1918 rows x 97 columns]
```

[25]:

```
# Notice that the results are still indexed, but that any of the results which␣
↪were in a group with a mean
# review score of less than or equal to 9.2 were not copied over.
```

2.4 Applying

[26]:
```
# By far the most common operation I invoke on groupby objects is the apply()␣
↪function. This allows you to
# apply an arbitrary function to each group, and stitch the results back for␣
↪each apply() into a single
# dataframe where the index is preserved.

# Lets look at an example using our airbnb data, I'm going to get a clean copy␣
↪of the dataframe
df=pd.read_csv("datasets/listings.csv")
# And lets just include some of the columns we were interested in previously
df=df[['cancellation_policy','review_scores_value']]
df.head()
```

[26]:
```
  cancellation_policy  review_scores_value
0            moderate                  NaN
1            moderate                  9.0
2            moderate                 10.0
3            moderate                 10.0
4            flexible                 10.0
```

[27]:
```
# In previous work we wanted to find the average review score of a listing and␣
↪its deviation from the group
# mean. This was a two step process, first we used transform() on the groupby␣
↪object and then we had to
# broadcast to create a new column. With apply() we could wrap this logic in␣
↪one place
def calc_mean_review_scores(group):
    # group is a dataframe just of whatever we have grouped by, e.g.␣
↪cancellation policy, so we can treat
    # this as the complete dataframe
    avg=np.nanmean(group["review_scores_value"])
    # now broadcast our formula and create a new column
    group["review_scores_mean"]=np.abs(avg-group["review_scores_value"])
    return group

# Now just apply this to the groups
df.groupby('cancellation_policy').apply(calc_mean_review_scores).head()
```

[27]:
```
  cancellation_policy  review_scores_value  review_scores_mean
0            moderate                  NaN                 NaN
1            moderate                  9.0            0.307398
2            moderate                 10.0            0.692602
3            moderate                 10.0            0.692602
4            flexible                 10.0            0.762579
```

[28]:
```
# Using apply can be slower than using some of the specialized functions,␣
↪especially agg(). But, if your
# dataframes are not huge, it's a solid general purpose approach
```

Groupby is a powerful and commonly used tool for data cleaning and data analysis. Once you have grouped the data by some category you have a dataframe of just those values and you can conduct aggregated analsyis on the segments that you are interested. The groupby() function follows a split-apply-combine approach - first the data is split into subgroups, then you can apply some transformation, filtering, or aggregation, then the results are combined automatically by pandas for us.

Scales

November 15, 2021

1 Scales

[1]:
```
# Let's bring in pandas as normal
import pandas as pd

# Heres an example. Lets create a dataframe of letter grades in descending␣
↪order. We can also set an index
# value and here we'll just make it some human judgement of how good a student␣
↪was, like "excellent" or "good"

df=pd.DataFrame(['A+', 'A', 'A-', 'B+', 'B', 'B-', 'C+', 'C', 'C-', 'D+', 'D'],
                index=['excellent', 'excellent', 'excellent', 'good', 'good',␣
↪'good',
                       'ok', 'ok', 'ok', 'poor', 'poor'],
               columns=["Grades"])
df
```

[1]:
```
          Grades
excellent     A+
excellent      A
excellent     A-
good          B+
good           B
good          B-
ok            C+
ok             C
ok            C-
poor          D+
poor           D
```

[2]:
```
# Now, if we check the datatype of this column, we see that it's just an␣
↪object, since we set string values
df.dtypes
```

[2]:
```
Grades    object
dtype: object
```

[3]:
```
# We can, however, tell pandas that we want to change the type to category,␣
↪using the astype() function
df["Grades"].astype("category").head()
```

[3]:
```
excellent    A+
excellent     A
excellent    A-
good         B+
good          B
Name: Grades, dtype: category
Categories (11, object): [A, A+, A-, B, ..., C+, C-, D, D+]
```

[4]:
```
# We see now that there are eleven categories, and pandas is aware of what␣
↪those categories are. More
# interesting though is that our data isn't just categorical, but that it's␣
↪ordered. That is, an A- comes
# after a B+, and B comes before a B+. We can tell pandas that the data is␣
↪ordered by first creating a new
# categorical data type with the list of the categories (in order) and the␣
↪ordered=True flag
my_categories=pd.CategoricalDtype(categories=['D', 'D+', 'C-', 'C', 'C+', 'B-',␣
↪'B', 'B+', 'A-', 'A', 'A+'],
                           ordered=True)
# then we can just pass this to the astype() function
grades=df["Grades"].astype(my_categories)
grades.head()
```

[4]:
```
excellent    A+
excellent     A
excellent    A-
good         B+
good          B
Name: Grades, dtype: category
Categories (11, object): [D < D+ < C- < C ... B+ < A- < A < A+]
```

[5]:
```
# Now we see that pandas is not only aware that there are 11 categories, but it␣
↪is also aware of the order of
# those categoreies. So, what can you do with this? Well because there is an␣
↪ordering this can help with
# comparisons and boolean masking. For instance, if we have a list of our␣
↪grades and we compare them to a C
# we see that the lexicographical comparison returns results we were not␣
↪intending.

df[df["Grades"]>"C"]
```

[5]:
```
     Grades
ok       C+
ok       C-
```

```
poor      D+
poor       D
```

[6]:
```
# So a C+ is great than a C, but a C- and D certainly are not. However, if we␣
↪broadcast over the dataframe
# which has the type set to an ordered categorical

grades[grades>"C"]
```

[6]:
```
excellent    A+
excellent     A
excellent    A-
good         B+
good          B
good         B-
ok           C+
Name: Grades, dtype: category
Categories (11, object): [D < D+ < C- < C ... B+ < A- < A < A+]
```

[7]:
```
# We see that the operator works as we would expect. We can then use a certain␣
↪set of mathematical operators,
# like minimum, maximum, etc., on the ordinal data.
```

[8]:
```
# Sometimes it is useful to represent categorical values as each being a column␣
↪with a true or a false as to
# whether the category applies. This is especially common in feature␣
↪extraction, which is a topic in the data
# mining course. Variables with a boolean value are typically called dummy␣
↪variables, and pandas has a built
# in function called get_dummies which will convert the values of a single␣
↪column into multiple columns of
# zeros and ones indicating the presence of the dummy variable. I rarely use␣
↪it, but when I do it's very
# handy.
```

[9]:
```
# Theres one more common scale-based operation Id like to talk about, and thats␣
↪on converting a scale from
# something that is on the interval or ratio scale, like a numeric grade, into␣
↪one which is categorical. Now,
# this might seem a bit counter intuitive to you, since you are losing␣
↪information about the value. But its
# commonly done in a couple of places. For instance, if you are visualizing the␣
↪frequencies of categories,
# this can be an extremely useful approach, and histograms are regularly used␣
↪with converted interval or ratio
# data. In addition, if youre using a machine learning classification approach␣
↪on data, you need to be using
```

```python
# categorical data, so reducing dimensionality may be useful just to apply a␣
↪given technique. Pandas has a
# function called cut which takes as an argument some array-like structure like␣
↪a column of a dataframe or a
# series. It also takes a number of bins to be used, and all bins are kept at␣
↪equal spacing.

# Lets go back to our census data for an example. We saw that we could group by␣
↪state, then aggregate to get a
# list of the average county size by state. If we further apply cut to this␣
↪with, say, ten bins, we can see
# the states listed as categoricals using the average county size.

# let's bring in numpy
import numpy as np

# Now we read in our dataset
df=pd.read_csv("datasets/census.csv")

# And we reduce this to country data
df=df[df['SUMLEV']==50]

# And for a few groups
df=df.set_index('STNAME').groupby(level=0)['CENSUS2010POP'].agg(np.average)

df.head()
```

[9]:
```
STNAME
Alabama        71339.343284
Alaska         24490.724138
Arizona       426134.466667
Arkansas       38878.906667
California    642309.586207
Name: CENSUS2010POP, dtype: float64
```

[10]:
```python
# Now if we just want to make "bins" of each of these, we can use cut()
pd.cut(df,10)
```

[10]:
```
STNAME
Alabama                   (11706.087, 75333.413]
Alaska                    (11706.087, 75333.413]
Arizona                 (390320.176, 453317.529]
Arkansas                  (11706.087, 75333.413]
California              (579312.234, 642309.586]
Colorado                 (75333.413, 138330.766]
Connecticut             (390320.176, 453317.529]
Delaware                (264325.471, 327322.823]
District of Columbia    (579312.234, 642309.586]
```

```
Florida                (264325.471, 327322.823]
Georgia                  (11706.087, 75333.413]
Hawaii                 (264325.471, 327322.823]
Idaho                    (11706.087, 75333.413]
Illinois                (75333.413, 138330.766]
Indiana                  (11706.087, 75333.413]
Iowa                     (11706.087, 75333.413]
Kansas                   (11706.087, 75333.413]
Kentucky                 (11706.087, 75333.413]
Louisiana                (11706.087, 75333.413]
Maine                   (75333.413, 138330.766]
Maryland               (201328.118, 264325.471]
Massachusetts          (453317.529, 516314.881]
Michigan                (75333.413, 138330.766]
Minnesota                (11706.087, 75333.413]
Mississippi              (11706.087, 75333.413]
Missouri                 (11706.087, 75333.413]
Montana                  (11706.087, 75333.413]
Nebraska                 (11706.087, 75333.413]
Nevada                 (138330.766, 201328.118]
New Hampshire           (75333.413, 138330.766]
New Jersey             (390320.176, 453317.529]
New Mexico               (11706.087, 75333.413]
New York               (264325.471, 327322.823]
North Carolina          (75333.413, 138330.766]
North Dakota             (11706.087, 75333.413]
Ohio                    (75333.413, 138330.766]
Oklahoma                 (11706.087, 75333.413]
Oregon                  (75333.413, 138330.766]
Pennsylvania           (138330.766, 201328.118]
Rhode Island           (201328.118, 264325.471]
South Carolina          (75333.413, 138330.766]
South Dakota             (11706.087, 75333.413]
Tennessee                (11706.087, 75333.413]
Texas                   (75333.413, 138330.766]
Utah                    (75333.413, 138330.766]
Vermont                  (11706.087, 75333.413]
Virginia                 (11706.087, 75333.413]
Washington             (138330.766, 201328.118]
West Virginia            (11706.087, 75333.413]
Wisconsin               (75333.413, 138330.766]
Wyoming                  (11706.087, 75333.413]
Name: CENSUS2010POP, dtype: category
Categories (10, interval[float64]): [(11706.087, 75333.413] < (75333.413,
138330.766] < (138330.766, 201328.118] < (201328.118, 264325.471] ...
(390320.176, 453317.529] < (453317.529, 516314.881] < (516314.881, 579312.234] <
(579312.234, 642309.586]]
```

[11]:
```
# Here we see that states like alabama and alaska fall into the same category,␣
↪while california and the
# disctrict of columbia fall in a very different category.

# Now, cutting is just one way to build categories from your data, and there␣
↪are many other methods. For
# instance, cut gives you interval data, where the spacing between each␣
↪category is equal sized. But sometimes
# you want to form categories based on frequency  you want the number of items␣
↪in each bin to the be the
# same, instead of the spacing between bins. It really depends on what the␣
↪shape of your data is, and what
# youre planning to do with it.
```

PivotTable_ed

November 15, 2021

A pivot table is a way of summarizing data in a DataFrame for a particular purpose. It makes heavy use of the aggregation function. A pivot table is itself a DataFrame, where the rows represent one variable that you're interested in, the columns another, and the cell's some aggregate value. A pivot table also tends to includes marginal values as well, which are the sums for each column and row. This allows you to be able to see the relationship between two variables at just a glance.

[1]:
```
# Lets take a look at pivot tables in pandas
import pandas as pd
import numpy as np
```

[2]:
```
# Here we have the Times Higher Education World University Ranking dataset,␣
↪which is one of the most
# influential university measures. Let's import the dataset and see what it␣
↪looks like
df = pd.read_csv('datasets/cwurData.csv')
df.head()
```

[2]:
```
   world_rank                            institution         country  \
0           1                     Harvard University             USA
1           2  Massachusetts Institute of Technology             USA
2           3                    Stanford University             USA
3           4                University of Cambridge  United Kingdom
4           5     California Institute of Technology             USA

   national_rank  quality_of_education  alumni_employment  quality_of_faculty  \
0              1                     7                  9                   1
1              2                     9                 17                   3
2              3                    17                 11                   5
3              1                    10                 24                   4
4              4                     2                 29                   7

   publications  influence  citations  broad_impact  patents   score  year
0             1          1          1           NaN        5  100.00  2012
1            12          4          4           NaN        1   91.67  2012
2             4          2          2           NaN       15   89.50  2012
3            16         16         11           NaN       50   86.17  2012
4            37         22         22           NaN       18   85.21  2012
```

[3]:
```python
# Here we can see each institution's rank, country, quality of education, other␣
↪metrics, and overall score.
# Let's say we want to create a new column called Rank_Level, where␣
↪institutions with world ranking 1-100 are
# categorized as first tier and those with world ranking 101 - 200 are second␣
↪tier, ranking 201 - 300 are
# third tier, after 301 is other top universities.

# Now, you actually already have enough knowledge to do this, so why don't you␣
↪pause the video and give it a
# try?

# Here's my solution, I'm going to create a function called create_category␣
↪which will operate on the first
# column in the dataframe, world_rank
def create_category(ranking):
    # Since the rank is just an integer, I'll just do a bunch of if/elif␣
↪statements
    if (ranking >= 1) & (ranking <= 100):
        return "First Tier Top Unversity"
    elif (ranking >= 101) & (ranking <= 200):
        return "Second Tier Top Unversity"
    elif (ranking >= 201) & (ranking <= 300):
        return "Third Tier Top Unversity"
    return "Other Top Unversity"

# Now we can apply this to a single column of data to create a new series
df['Rank_Level'] = df['world_rank'].apply(lambda x: create_category(x))
# And lets look at the result
df.head()
```

[3]:
```
   world_rank                            institution         country  \
0           1                     Harvard University             USA
1           2  Massachusetts Institute of Technology             USA
2           3                    Stanford University             USA
3           4               University of Cambridge  United Kingdom
4           5     California Institute of Technology             USA

   national_rank  quality_of_education  alumni_employment  quality_of_faculty  \
0              1                     7                  9                   1
1              2                     9                 17                   3
2              3                    17                 11                   5
3              1                    10                 24                   4
4              4                     2                 29                   7

   publications  influence  citations  broad_impact  patents   score  year  \
0             1          1          1           NaN        5  100.00  2012
```

```
1            12         4         4          NaN       1   91.67  2012
2             4         2         2          NaN      15   89.50  2012
3            16        16        11          NaN      50   86.17  2012
4            37        22        22          NaN      18   85.21  2012

                Rank_Level
0  First Tier Top Unversity
1  First Tier Top Unversity
2  First Tier Top Unversity
3  First Tier Top Unversity
4  First Tier Top Unversity
```

[4]:

```
# A pivot table allows us to pivot out one of these columns a new column␣
 ↪headers and compare it against
# another column as row indices. Let's say we want to compare rank level versus␣
 ↪country of the universities
# and we want to compare in terms of overall score

# To do this, we tell Pandas we want the values to be Score, and index to be␣
 ↪the country and the columns to be
# the rank levels. Then we specify that the aggregation function, and here␣
 ↪we'll use the NumPy mean to get the
# average rating for universities in that country

df.pivot_table(values='score', index='country', columns='Rank_Level',␣
 ↪aggfunc=[np.mean]).head()
```

[4]:

```
                                mean                      \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                        NaN           44.672857
Australia                    47.9425           44.645750
Austria                          NaN           44.864286
Belgium                      51.8750           45.081000
Brazil                           NaN           44.499706


Rank_Level Second Tier Top Unversity Third Tier Top Unversity
country
Argentina                         NaN                      NaN
Australia                     49.2425                47.285000
Austria                           NaN                47.066667
Belgium                       49.0840                46.746667
Brazil                        49.5650                      NaN
```

[5]:

```
# We can see a  hierarchical dataframe where the index, or rows, are by country␣
 ↪and the columns have two
```

```
# levels, the top level indicating that the mean value is being used and the
↪second level being our ranks. In
# this example we only have one variable, the mean, that we are looking at, so
↪we don't really need a
# heirarchical index.

# We notice that there are some NaN values, for example, the first row,
↪Argentia. The NaN values indicate that
# Argentia has only observations in the "Other Top Unversities" category
```

[6]:

```
# Now, pivot tables aren't limited to one function that you might want to apply.
↪ You can pass a named
# parameter, aggfunc, which is a list of the different functions to apply, and
↪pandas will provide you with
# the result using hierarchical column names.  Let's try that same query, but
↪pass in the max() function too

df.pivot_table(values='score', index='country', columns='Rank_Level',
↪aggfunc=[np.mean, np.max]).head()
```

[6]:

```
                               mean                           \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                       NaN           44.672857
Australia                   47.9425           44.645750
Austria                         NaN           44.864286
Belgium                     51.8750           45.081000
Brazil                          NaN           44.499706


                                                                   \
Rank_Level Second Tier Top Unversity Third Tier Top Unversity
country
Argentina                        NaN                      NaN
Australia                    49.2425                47.285000
Austria                          NaN                47.066667
Belgium                      49.0840                46.746667
Brazil                       49.5650                      NaN

                               amax                           \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                       NaN               45.66
Australia                     51.61               45.97
Austria                         NaN               46.29
Belgium                       52.03               46.21
Brazil                          NaN               46.08
```

```
Rank_Level Second Tier Top Unversity Third Tier Top Unversity
country
Argentina                        NaN                      NaN
Australia                      50.40                    47.47
Austria                          NaN                    47.78
Belgium                        49.73                    47.14
Brazil                         49.82                      NaN
```

[7]:
```python
# So now we see we have both the mean and the max. As mentioned earlier, we can␣
↪also summarize the values
# within a given top level column. For instance, if we want to see an overall␣
↪average for the country for the
# mean and we want to see the max of the max, we can indicate that we want␣
↪pandas to provide marginal values
df.pivot_table(values='score', index='country', columns='Rank_Level',␣
↪aggfunc=[np.mean, np.max],
               margins=True).head()
```

[7]:
```
                               mean                      \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                       NaN           44.672857
Australia                   47.9425           44.645750
Austria                         NaN           44.864286
Belgium                     51.8750           45.081000
Brazil                          NaN           44.499706

                                                                   \
Rank_Level Second Tier Top Unversity Third Tier Top Unversity        All
country
Argentina                        NaN                      NaN  44.672857
Australia                    49.2425                47.285000  45.825517
Austria                          NaN                47.066667  45.139583
Belgium                      49.0840                46.746667  47.011000
Brazil                       49.5650                      NaN  44.781111

                               amax                      \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                       NaN               45.66
Australia                     51.61               45.97
Austria                         NaN               46.29
Belgium                       52.03               46.21
Brazil                          NaN               46.08


Rank_Level Second Tier Top Unversity Third Tier Top Unversity    All
```

```
country
Argentina                          NaN                        NaN  45.66
Australia                        50.40                      47.47  51.61
Austria                            NaN                      47.78  47.78
Belgium                          49.73                      47.14  52.03
Brazil                           49.82                        NaN  49.82
```

[8]:
```
# A pivot table is just a multi-level dataframe, and we can access series or␣
↪cells in the dataframe in a similar way
# as we do so for a regular dataframe.

# Let's create a new dataframe from our previous example
new_df=df.pivot_table(values='score', index='country', columns='Rank_Level',␣
↪aggfunc=[np.mean, np.max],
               margins=True)
# Now let's look at the index
print(new_df.index)
# And let's look at the columns
print(new_df.columns)
```

```
Index(['Argentina', 'Australia', 'Austria', 'Belgium', 'Brazil', 'Bulgaria',
       'Canada', 'Chile', 'China', 'Colombia', 'Croatia', 'Cyprus',
       'Czech Republic', 'Denmark', 'Egypt', 'Estonia', 'Finland', 'France',
       'Germany', 'Greece', 'Hong Kong', 'Hungary', 'Iceland', 'India', 'Iran',
       'Ireland', 'Israel', 'Italy', 'Japan', 'Lebanon', 'Lithuania',
       'Malaysia', 'Mexico', 'Netherlands', 'New Zealand', 'Norway', 'Poland',
       'Portugal', 'Puerto Rico', 'Romania', 'Russia', 'Saudi Arabia',
       'Serbia', 'Singapore', 'Slovak Republic', 'Slovenia', 'South Africa',
       'South Korea', 'Spain', 'Sweden', 'Switzerland', 'Taiwan', 'Thailand',
       'Turkey', 'USA', 'Uganda', 'United Arab Emirates', 'United Kingdom',
       'Uruguay', 'All'],
      dtype='object', name='country')
MultiIndex([('mean',  'First Tier Top Unversity'),
            ('mean',       'Other Top Unversity'),
            ('mean', 'Second Tier Top Unversity'),
            ('mean',  'Third Tier Top Unversity'),
            ('mean',                        'All'),
            ('amax',  'First Tier Top Unversity'),
            ('amax',       'Other Top Unversity'),
            ('amax', 'Second Tier Top Unversity'),
            ('amax',  'Third Tier Top Unversity'),
            ('amax',                        'All')],
           names=[None, 'Rank_Level'])
```

[9]:
```
# We can see the columns are hierarchical. The top level column indices have␣
↪two categories: mean and max, and
```

```
# the lower level column indices have four categories, which are the four rank␣
↪levels. How would we query this
# if we want to get the average scores of First Tier Top Unversity levels in␣
↪each country? We would just need
# to make two dataframe projections, the first for the mean, then the second␣
↪for the top tier
new_df['mean']['First Tier Top Unversity'].head()
```

[9]:
```
country
Argentina        NaN
Australia    47.9425
Austria          NaN
Belgium      51.8750
Brazil           NaN
Name: First Tier Top Unversity, dtype: float64
```

[10]:
```
# We can see that the output is a series object which we can confirm by␣
↪printing the type. Remember that when
# you project a single column of values out of a DataFrame you get a series.
type(new_df['mean']['First Tier Top Unversity'])
```

[10]:
```
pandas.core.series.Series
```

[11]:
```
# What if we want to find the country that has the maximum average score on␣
↪First Tier Top University level?
# We can use the idxmax() function.
new_df['mean']['First Tier Top Unversity'].idxmax()
```

[11]:
```
'United Kingdom'
```

[12]:
```
# Now, the idxmax() function isn't special for pivot tables, it's a built in␣
↪function to the Series object.
# We don't have time to go over all pandas functions and attributes, and I want␣
↪to encourage you to explore
# the API to learn more deeply what is available to you.
```

[13]:
```
# If you want to achieve a different shape of your pivot table, you can do so␣
↪with the stack and unstack
# functions. Stacking is pivoting the lowermost column index to become the␣
↪innermost row index. Unstacking is
# the inverse of stacking, pivoting the innermost row index to become the␣
↪lowermost column index. An example
# will help make this clear

# Let's look at our pivot table first to refresh what it looks like
new_df.head()
```

[13]:
```
                                 mean                          \
Rank_Level First Tier Top Unversity Other Top Unversity
country
```

```
Argentina                     NaN          44.672857
Australia                  47.9425         44.645750
Austria                       NaN          44.864286
Belgium                    51.8750         45.081000
Brazil                        NaN          44.499706

                                                                          \
Rank_Level Second Tier Top Unversity Third Tier Top Unversity        All
country
Argentina                      NaN                      NaN  44.672857
Australia                  49.2425                47.285000  45.825517
Austria                        NaN                47.066667  45.139583
Belgium                    49.0840                46.746667  47.011000
Brazil                     49.5650                      NaN  44.781111

                              amax                          \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                     NaN               45.66
Australia                   51.61               45.97
Austria                       NaN               46.29
Belgium                     52.03               46.21
Brazil                        NaN               46.08


Rank_Level Second Tier Top Unversity Third Tier Top Unversity    All
country
Argentina                      NaN                      NaN  45.66
Australia                    50.40                    47.47  51.61
Austria                        NaN                    47.78  47.78
Belgium                      49.73                    47.14  52.03
Brazil                       49.82                      NaN  49.82
```

[14]:
```
# Now let's try stacking, this should move the lowermost column, so the tiers␣
↪of the university rankings, to
# the inner most row
new_df=new_df.stack()
new_df.head()
```

[14]:
```
                                           mean   amax
country   Rank_Level
Argentina Other Top Unversity         44.672857  45.66
          All                         44.672857  45.66
Australia First Tier Top Unversity    47.942500  51.61
          Other Top Unversity         44.645750  45.97
          Second Tier Top Unversity   49.242500  50.40
```

[15]:
```
# In the original pivot table, rank levels are the lowermost column, after␣
↪stacking, rank levels become the
# innermost index, appearing to the right after country

# Now let's try unstacking
new_df.unstack().head()
```

[15]:
```
                                    mean                      \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                            NaN           44.672857
Australia                        47.9425           44.645750
Austria                              NaN           44.864286
Belgium                          51.8750           45.081000
Brazil                               NaN           44.499706

                                                                      \
Rank_Level Second Tier Top Unversity Third Tier Top Unversity        All
country
Argentina                             NaN                      NaN  44.672857
Australia                         49.2425                47.285000  45.825517
Austria                               NaN                47.066667  45.139583
Belgium                           49.0840                46.746667  47.011000
Brazil                            49.5650                      NaN  44.781111

                                    amax                      \
Rank_Level First Tier Top Unversity Other Top Unversity
country
Argentina                            NaN               45.66
Australia                          51.61               45.97
Austria                              NaN               46.29
Belgium                            52.03               46.21
Brazil                               NaN               46.08


Rank_Level Second Tier Top Unversity Third Tier Top Unversity    All
country
Argentina                             NaN                      NaN  45.66
Australia                           50.40                    47.47  51.61
Austria                               NaN                    47.78  47.78
Belgium                             49.73                    47.14  52.03
Brazil                              49.82                      NaN  49.82
```

[16]:
```
# That seems to restore our dataframe to its original shape. What do you think␣
↪would happen if we unstacked twice in a row?
new_df.unstack().unstack().head()
```

[16]:

```
      Rank_Level                 country
mean  First Tier Top Unversity   Argentina        NaN
                                 Australia    47.9425
                                 Austria          NaN
                                 Belgium      51.8750
                                 Brazil           NaN
dtype: float64
```

[17]:

```
# We actually end up unstacking all the way to just a single column, so a␣
 ↪series object is returned. This
# column is just a "value", the meaning of which is denoted by the␣
 ↪heirarachical index of operation, rank, and
# country.
```

So that's pivot tables. This has been a pretty short description, but they're incredibly useful when dealing with numeric data, especially if you're trying to summarize the data in some form. You'll regularly be creating new pivot tables on slices of data, whether you're exploring the data yourself or preparing data for others to report on. And of course, you can pass any function you want to the aggregate function, including those that you define yourself.

DateFunctionality_ed

November 15, 2021

In today's lecture, where we'll be looking at the time series and date functionally in pandas. Manipulating dates and time is quite flexible in Pandas and thus allows us to conduct more analysis such as time series analysis, which we will talk about soon. Actually, pandas was originally created by Wed McKinney to handle date and time data when he worked as a consultant for hedge funds.

[1]:
```python
# Let's bring in pandas and numpy as usual
import pandas as pd
import numpy as np
```

0.0.1 Timestamp

[2]:
```python
# Pandas has four main time related classes. Timestamp, DatetimeIndex, Period,␣
↪and PeriodIndex. First, let's
# look at Timestamp. It represents a single timestamp and associates values␣
↪with points in time.

# For example, let's create a timestamp using a string 9/1/2019 10:05AM, and␣
↪here we have our timestamp.
# Timestamp is interchangeable with Python's datetime in most cases.
pd.Timestamp('9/1/2019 10:05AM')
```

[2]:
```
Timestamp('2019-09-01 10:05:00')
```

[3]:
```python
# We can also create a timestamp by passing multiple parameters such as year,␣
↪month, date, hour,
# minute, separately
pd.Timestamp(2019, 12, 20, 0, 0)
```

[3]:
```
Timestamp('2019-12-20 00:00:00')
```

[4]:
```python
# Timestamp also has some useful attributes, such as isoweekday(), which shows␣
↪the weekday of the timestamp
# note that 1 represents Monday and 7 represents Sunday
pd.Timestamp(2019, 12, 20, 0, 0).isoweekday()
```

[4]:
```
5
```

[5]:
```python
# You can find extract the specific year, month, day, hour, minute, second from␣
↪a timestamp
```

```
pd.Timestamp(2019, 12, 20, 5, 2,23).second
```

[5]: 23

0.0.2 Period

[6]:
```
# Suppose we weren't interested in a specific point in time and instead wanted␣
↪a span of time. This is where
# the Period class comes into play. Period represents a single time span, such␣
↪as a specific day or month.

# Here we are creating a period that is January 2016,
pd.Period('1/2016')
```

[6]: Period('2016-01', 'M')

[7]:
```
# You'll notice when we print that out that the granularity of the period is M␣
↪for month, since that was the
# finest grained piece we provided. Here's an example of a period that is March␣
↪5th, 2016.
pd.Period('3/5/2016')
```

[7]: Period('2016-03-05', 'D')

[8]:
```
# Period objects represent the full timespan that you specify. Arithmetic on␣
↪period is very easy and
# intuitive, for instance, if we want to find out 5 months after January 2016,␣
↪we simply plus 5
pd.Period('1/2016') + 5
```

[8]: Period('2016-06', 'M')

[9]:
```
# From the result, you can see we get June 2016. If we want to find out two␣
↪days before March 5th 2016, we
# simply subtract 2
pd.Period('3/5/2016') - 2
```

[9]: Period('2016-03-03', 'D')

[10]:
```
# The key here is that the period object encapsulates the granularity for␣
↪arithmetic
```

0.0.3 DatetimeIndex and PeriodIndex

[11]:
```
# The index of a timestamp is DatetimeIndex. Let's look at a quick example.␣
↪First, let's create our example
# series t1, we'll use the Timestamp of September 1st, 2nd and 3rd of 2016.␣
↪When we look at the series, each
# Timestamp is the index and has a value associated with it, in this case, a, b␣
↪and c.
```

```
t1 = pd.Series(list('abc'), [pd.Timestamp('2016-09-01'), pd.
 ↪Timestamp('2016-09-02'),
                             pd.Timestamp('2016-09-03')])
t1
```

[11]:
```
2016-09-01    a
2016-09-02    b
2016-09-03    c
dtype: object
```

[12]:
```
# Looking at the type of our series index, we see that it's DatetimeIndex.
type(t1.index)
```

[12]:
```
pandas.core.indexes.datetimes.DatetimeIndex
```

[13]:
```
# Similarly, we can create a period-based index as well.
t2 = pd.Series(list('def'), [pd.Period('2016-09'), pd.Period('2016-10'),
                             pd.Period('2016-11')])
t2
```

[13]:
```
2016-09    d
2016-10    e
2016-11    f
Freq: M, dtype: object
```

[14]:
```
# Looking at the type of the ts2.index, we can see that it's PeriodIndex.
type(t2.index)
```

[14]:
```
pandas.core.indexes.period.PeriodIndex
```

0.0.4 Converting to Datetime

[15]:
```
# Now, let's look into how to convert to Datetime. Suppose we have a list of␣
 ↪dates as strings and we want to
# create a new dataframe

# I'm going to try a bunch of different date formats
d1 = ['2 June 2013', 'Aug 29, 2014', '2015-06-26', '7/12/16']

# And just some random data
ts3 = pd.DataFrame(np.random.randint(10, 100, (4,2)), index=d1,
                   columns=list('ab'))
ts3
```

[15]:

	a	b
2 June 2013	38	40
Aug 29, 2014	94	99
2015-06-26	75	32
7/12/16	68	62

[16]:
```
# Using pandas to_datetime, pandas will try to convert these to Datetime and␣
↪put them in a standard format.

ts3.index = pd.to_datetime(ts3.index)
ts3
```

[16]:
```
             a   b
2013-06-02  38  40
2014-08-29  94  99
2015-06-26  75  32
2016-07-12  68  62
```

[17]:
```
# to_datetime also() has options to change the date parse order. For example,␣
↪we
# can pass in the argument dayfirst = True to parse the date in European date.

pd.to_datetime('4.7.12', dayfirst=True)
```

[17]:
```
Timestamp('2012-07-04 00:00:00')
```

0.0.5 Timedelta

[18]:
```
# Timedeltas are differences in times. This is not the same as a a period, but␣
↪conceptually similar. For
# instance, if we want to take the difference between September 3rd and ␣
↪September 1st, we get a Timedelta of
# two days.
pd.Timestamp('9/3/2016')-pd.Timestamp('9/1/2016')
```

[18]:
```
Timedelta('2 days 00:00:00')
```

[19]:
```
# We can also do something like find what the date and time is for 12 days and␣
↪three hours past September 2nd,
# at 8:10 AM.
pd.Timestamp('9/2/2016 8:10AM') + pd.Timedelta('12D 3H')
```

[19]:
```
Timestamp('2016-09-14 11:10:00')
```

0.0.6 Offset

[20]:
```
# Offset is similar to timedelta, but it follows specific calendar duration␣
↪rules. Offset allows flexibility
# in terms of types of time intervals. Besides hour, day, week, month, etc it␣
↪also has business day, end of
# month, semi month begin etc

# Let's create a timestamp, and see what day is that
pd.Timestamp('9/4/2016').weekday()
```

[20]: 6

[21]:
```
# Now we can now add the timestamp with a week ahead
pd.Timestamp('9/4/2016') + pd.offsets.Week()
```

[21]: `Timestamp('2016-09-11 00:00:00')`

[22]:
```
# Now let's try to do the month end, then we would have the last day of␣
↪Septemer
pd.Timestamp('9/4/2016') + pd.offsets.MonthEnd()
```

[22]: `Timestamp('2016-09-30 00:00:00')`

0.0.7 Working with Dates in a Dataframe

[23]:
```
# Next, let's look at a few tricks for working with dates in a DataFrame.␣
↪Suppose we want to look at nine
# measurements, taken bi-weekly, every Sunday, starting in October 2016. Using␣
↪date_range, we can create this
# DatetimeIndex. In data_range, we have to either specify the start or end date.
↪ If it is not explicitly
# specified, by default, the date is considered the start date. Then we have to␣
↪specify number of periods, and
# a frequency. Here, we set it to "2W-SUN", which means biweekly on Sunday

dates = pd.date_range('10-01-2016', periods=9, freq='2W-SUN')
dates
```

[23]:
```
DatetimeIndex(['2016-10-02', '2016-10-16', '2016-10-30', '2016-11-13',
               '2016-11-27', '2016-12-11', '2016-12-25', '2017-01-08',
               '2017-01-22'],
              dtype='datetime64[ns]', freq='2W-SUN')
```

[24]:
```
# There are many other frequencies that you can specify. For example, you can␣
↪do business day
pd.date_range('10-01-2016', periods=9, freq='B')
```

[24]:
```
DatetimeIndex(['2016-10-03', '2016-10-04', '2016-10-05', '2016-10-06',
               '2016-10-07', '2016-10-10', '2016-10-11', '2016-10-12',
               '2016-10-13'],
              dtype='datetime64[ns]', freq='B')
```

[25]:
```
# Or you can do quarterly, with the quarter start in June
pd.date_range('04-01-2016', periods=12, freq='QS-JUN')
```

[25]:
```
DatetimeIndex(['2016-06-01', '2016-09-01', '2016-12-01', '2017-03-01',
               '2017-06-01', '2017-09-01', '2017-12-01', '2018-03-01',
               '2018-06-01', '2018-09-01', '2018-12-01', '2019-03-01'],
              dtype='datetime64[ns]', freq='QS-JUN')
```

[26]:
```
# Now, let's go back to our weekly on Sunday example and create a DataFrame␣
↪using these dates, and some random
# data, and see what we can do with it.

dates = pd.date_range('10-01-2016', periods=9, freq='2W-SUN')
df = pd.DataFrame({'Count 1': 100 + np.random.randint(-5, 10, 9).cumsum(),
                  'Count 2': 120 + np.random.randint(-5, 10, 9)}, index=dates)
df
```

[26]:

	Count 1	Count 2
2016-10-02	109	119
2016-10-16	109	127
2016-10-30	117	118
2016-11-13	115	129
2016-11-27	113	122
2016-12-11	122	129
2016-12-25	127	129
2017-01-08	135	117
2017-01-22	138	128

[27]:
```
# First, we can check what day of the week a specific date is. For example,␣
↪here we can see that all the dates
# in our index are on a Sunday. Which matches the frequency that we set
df.index.weekday_name
```

[27]:
```
Index(['Sunday', 'Sunday', 'Sunday', 'Sunday', 'Sunday', 'Sunday', 'Sunday',
       'Sunday', 'Sunday'],
      dtype='object')
```

[28]:
```
# We can also use diff() to find the difference between each date's value.
df.diff()
```

[28]:

	Count 1	Count 2
2016-10-02	NaN	NaN
2016-10-16	0.0	8.0
2016-10-30	8.0	-9.0
2016-11-13	-2.0	11.0
2016-11-27	-2.0	-7.0
2016-12-11	9.0	7.0
2016-12-25	5.0	0.0
2017-01-08	8.0	-12.0
2017-01-22	3.0	11.0

[29]:
```
# Suppose we want to know what the mean count is for each month in our␣
↪DataFrame. We can do this using
# resample. Converting from a higher frequency from a lower frequency is called␣
↪downsampling (we'll talk about
# this in a moment)
df.resample('M').mean()
```

[29]:
```
                Count 1     Count 2
2016-10-31  111.666667  121.333333
2016-11-30  114.000000  125.500000
2016-12-31  124.500000  129.000000
2017-01-31  136.500000  122.500000
```

[30]:
```
# Now let's talk about datetime indexing and slicing, which is a wonderful
 ↪feature of the pandas DataFrame.
# For instance, we can use partial string indexing to find values from a
 ↪particular year,
df['2017']
```

[30]:
```
            Count 1  Count 2
2017-01-08      135      117
2017-01-22      138      128
```

[31]:
```
# Or we can do it from a particular month
df['2016-12']
```

[31]:
```
            Count 1  Count 2
2016-12-11      122      129
2016-12-25      127      129
```

[32]:
```
# Or we can even slice on a range of dates For example, here we only want the
 ↪values from December 2016
# onwards.
df['2016-12':]
```

[32]:
```
            Count 1  Count 2
2016-12-11      122      129
2016-12-25      127      129
2017-01-08      135      117
2017-01-22      138      128
```

[33]:
```
df['2016']
```

[33]:
```
            Count 1  Count 2
2016-10-02      109      119
2016-10-16      109      127
2016-10-30      117      118
2016-11-13      115      129
2016-11-27      113      122
2016-12-11      122      129
2016-12-25      127      129
```

assignment3

November 15, 2021

1 Assignment 3

All questions are weighted the same in this assignment. This assignment requires more individual learning then the last one did - you are encouraged to check out the pandas documentation to find functions or methods you might not have used yet, or ask questions on Stack Overflow and tag them as pandas and python related. All questions are worth the same number of points except question 1 which is worth 17% of the assignment grade.

Note: Questions 3-13 rely on your question 1 answer.

[27]:
```
import pandas as pd
import numpy as np

# Filter all warnings. If you would like to see the warnings, please comment␣
↪the two lines below.
import warnings
warnings.filterwarnings('ignore')
```

1.0.1 Question 1

Load the energy data from the file `assets/Energy Indicators.xls`, which is a list of indicators of energy supply and renewable electricity production from the United Nations for the year 2013, and should be put into a DataFrame with the variable name of **Energy**.

Keep in mind that this is an Excel file, and not a comma separated values file. Also, make sure to exclude the footer and header information from the datafile. The first two columns are unneccessary, so you should get rid of them, and you should change the column labels so that the columns are:

`['Country', 'Energy Supply', 'Energy Supply per Capita', '% Renewable]`

Convert `Energy Supply` to gigajoules (**Note: there are 1,000,000 gigajoules in a petajoule**). For all countries which have missing data (e.g. data with "...") make sure this is reflected as `np.NaN` values.

Rename the following list of countries (for use in later questions):

`"Republic of Korea": "South Korea", "United States of America": "United States", "United Kingdom of Great Britain and Northern Ireland": "United Kingdom", "China, Hong Kong Special Administrative Region": "Hong Kong"`

There are also several countries with numbers and/or parenthesis in their name. Be sure to remove these, e.g. `'Bolivia (Plurinational State of)'` should be `'Bolivia'`. `'Switzerland17'` should be `'Switzerland'`.

Next, load the GDP data from the file `assets/world_bank.csv`, which is a csv containing countries' GDP from 1960 to 2015 from World Bank. Call this DataFrame **GDP**.

Make sure to skip the header, and rename the following list of countries:

`"Korea, Rep.": "South Korea", "Iran, Islamic Rep.": "Iran", "Hong Kong SAR, China": "Hong Kong"`

Finally, load the Sciamgo Journal and Country Rank data for Energy Engineering and Power Technology from the file `assets/scimagojr-3.xlsx`, which ranks countries based on their journal contributions in the aforementioned area. Call this DataFrame **ScimEn**.

Join the three datasets: GDP, Energy, and ScimEn into a new dataset (using the intersection of country names). Use only the last 10 years (2006-2015) of GDP data and only the top 15 countries by Scimagojr 'Rank' (Rank 1 through 15).

The index of this DataFrame should be the name of the country, and the columns should be ['Rank', 'Documents', 'Citable documents', 'Citations', 'Self-citations', 'Citations per document', 'H index', 'Energy Supply', 'Energy Supply per Capita', '% Renewable', '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015'].

This function should return a DataFrame with 20 columns and 15 entries, and the rows of the DataFrame should be sorted by "Rank".

[24]:
```
def answer_one():
    # YOUR CODE HERE
    import pandas as pd
    import numpy as np
    import re
    Energy=pd.read_excel('assets/Energy Indicators.xls',header=1,skipfooter=1)
    Energy=Energy.drop(['Unnamed: 0','Unnamed: 1'],axis=1)
    for i in range(243,280):
        Energy=Energy.drop(i)
    for i in range(0,16):
        Energy=Energy.drop(i)
    Energy.columns=['Country','Energy Supply','Energy Supply per Capita', '%␣
↪Renewable']
    Energy['Energy Supply']=Energy['Energy Supply']*1000000
    Energy['Energy Supply']=Energy['Energy Supply'].replace('...',np.nan)

    Energy['Country']=Energy['Country'].replace("Bolivia (Plurinational State␣
↪of)","Bolivia")
    pattern2="""(?P<Country2>.*?)(?P<bidule>\s\(.*\)|)"""
    pattern1="""(?P<Country2>.*?)(?P<Num>\d*$)"""
    extract1=Energy['Country'].str.extract(pattern1)
    Energy['Country']=extract1['Country2']
    Energy['Energy Supply per Capita']=Energy['Energy Supply per Capita'].
↪replace('...',np.nan)
    Energy['Country']=Energy['Country'].replace("Republic of Korea", "South␣
↪Korea")
    Energy['Country']=Energy['Country'].replace("United States of America",␣
↪"United States")
```

```
    Energy['Country']=Energy['Country'].replace("United Kingdom of Great␣
↪Britain and Northern Ireland", "United Kingdom")
    Energy['Country']=Energy['Country'].replace("China, Hong Kong Special␣
↪Administrative Region", "Hong Kong")
    Energy['Country']=Energy['Country'].replace("Iran (Islamic Republic of)",␣
↪"Iran")

    #extract2=Energy['Country'].str.extract(pattern2)
    #Energy['Country']=extract2['Country2']
    GDP=pd.read_csv('assets/world_bank.csv',header=4)

    GDP=GDP.rename({'Country Name':'Country'},axis='columns')
    GDP['Country']=GDP['Country'].replace({"Korea, Rep.":"South Korea", "Iran,␣
↪Islamic Rep.":"Iran","Hong Kong SAR, China":"Hong Kong"})
  ␣
↪GDPmerge=GDP[['Country','2006','2007','2008','2009','2010','2011','2012','2013','2014','201
    ScimEn=pd.read_excel('assets/scimagojr-3.xlsx')
    ScimEnmerge=ScimEn[ScimEn['Rank']<16]
    #print(len(ScimEn)-len(ScimEnmerge))
    merge1=pd.merge(ScimEnmerge,Energy,how='left',on='Country')
    merge2=pd.merge(merge1,GDPmerge,how='left',on='Country')
    merge2=merge2.set_index('Country')
    return merge2
answer_one()
```

[24]:

```
                    Rank  Documents  Citable documents  Citations  \
Country
China                  1     127050             126767     597237
United States          2      96661              94747     792274
Japan                  3      30504              30287     223024
United Kingdom         4      20944              20357     206091
Russian Federation     5      18534              18301      34266
Canada                 6      17899              17620     215003
Germany                7      17027              16831     140566
India                  8      15005              14841     128763
France                 9      13153              12973     130632
South Korea           10      11983              11923     114675
Italy                 11      10964              10794     111850
Spain                 12       9428               9330     123336
Iran                  13       8896               8819      57470
Australia             14       8831               8725      90765
Brazil                15       8668               8596      60702

                    Self-citations  Citations per document  H index  \
Country
China                       411683                    4.70      138
United States               265436                    8.20      230
```

Japan	61554	7.31	134
United Kingdom	37874	9.84	139
Russian Federation	12422	1.85	57
Canada	40930	12.01	149
Germany	27426	8.26	126
India	37209	8.58	115
France	28601	9.93	114
South Korea	22595	9.57	104
Italy	26661	10.20	106
Spain	23964	13.08	115
Iran	19125	6.46	72
Australia	15606	10.28	107
Brazil	14396	7.00	86

	Energy Supply	Energy Supply per Capita	% Renewable \
Country			
China	127191000000	93.0	19.7549
United States	90838000000	286.0	11.571
Japan	18984000000	149.0	10.2328
United Kingdom	7920000000	124.0	10.6005
Russian Federation	30709000000	214.0	17.2887
Canada	10431000000	296.0	61.9454
Germany	13261000000	165.0	17.9015
India	33195000000	26.0	14.9691
France	10597000000	166.0	17.0203
South Korea	11007000000	221.0	2.27935
Italy	6530000000	109.0	33.6672
Spain	4923000000	106.0	37.9686
Iran	9172000000	119.0	5.70772
Australia	5386000000	231.0	11.8108
Brazil	12149000000	59.0	69.648

	2006	2007	2008	2009 \
Country				
China	3.992331e+12	4.559041e+12	4.997775e+12	5.459247e+12
United States	1.479230e+13	1.505540e+13	1.501149e+13	1.459484e+13
Japan	5.496542e+12	5.617036e+12	5.558527e+12	5.251308e+12
United Kingdom	2.419631e+12	2.482203e+12	2.470614e+12	2.367048e+12
Russian Federation	1.385793e+12	1.504071e+12	1.583004e+12	1.459199e+12
Canada	1.564469e+12	1.596740e+12	1.612713e+12	1.565145e+12
Germany	3.332891e+12	3.441561e+12	3.478809e+12	3.283340e+12
India	1.265894e+12	1.374865e+12	1.428361e+12	1.549483e+12
France	2.607840e+12	2.669424e+12	2.674637e+12	2.595967e+12
South Korea	9.410199e+11	9.924316e+11	1.020510e+12	1.027730e+12
Italy	2.202170e+12	2.234627e+12	2.211154e+12	2.089938e+12
Spain	1.414823e+12	1.468146e+12	1.484530e+12	1.431475e+12
Iran	3.895523e+11	4.250646e+11	4.289909e+11	4.389208e+11

```
Australia           1.021939e+12  1.060340e+12  1.099644e+12  1.119654e+12
Brazil              1.845080e+12  1.957118e+12  2.056809e+12  2.054215e+12

                            2010          2011          2012          2013  \
Country
China               6.039659e+12  6.612490e+12  7.124978e+12  7.672448e+12
United States       1.496437e+13  1.520402e+13  1.554216e+13  1.577367e+13
Japan               5.498718e+12  5.473738e+12  5.569102e+12  5.644659e+12
United Kingdom      2.403504e+12  2.450911e+12  2.479809e+12  2.533370e+12
Russian Federation  1.524917e+12  1.589943e+12  1.645876e+12  1.666934e+12
Canada              1.613406e+12  1.664087e+12  1.693133e+12  1.730688e+12
Germany             3.417298e+12  3.542371e+12  3.556724e+12  3.567317e+12
India               1.708459e+12  1.821872e+12  1.924235e+12  2.051982e+12
France              2.646995e+12  2.702032e+12  2.706968e+12  2.722567e+12
South Korea         1.094499e+12  1.134796e+12  1.160809e+12  1.194429e+12
Italy               2.125185e+12  2.137439e+12  2.077184e+12  2.040871e+12
Spain               1.431673e+12  1.417355e+12  1.380216e+12  1.357139e+12
Iran                4.677902e+11  4.853309e+11  4.532569e+11  4.445926e+11
Australia           1.142251e+12  1.169431e+12  1.211913e+12  1.241484e+12
Brazil              2.208872e+12  2.295245e+12  2.339209e+12  2.409740e+12

                            2014          2015
Country
China               8.230121e+12  8.797999e+12
United States       1.615662e+13  1.654857e+13
Japan               5.642884e+12  5.669563e+12
United Kingdom      2.605643e+12  2.666333e+12
Russian Federation  1.678709e+12  1.616149e+12
Canada              1.773486e+12  1.792609e+12
Germany             3.624386e+12  3.685556e+12
India               2.200617e+12  2.367206e+12
France              2.729632e+12  2.761185e+12
South Korea         1.234340e+12  1.266580e+12
Italy               2.033868e+12  2.049316e+12
Spain               1.375605e+12  1.419821e+12
Iran                4.639027e+11           NaN
Australia           1.272520e+12  1.301251e+12
Brazil              2.412231e+12  2.319423e+12
```

[2]:

```
assert type(answer_one()) == pd.DataFrame, "Q1: You should return a DataFrame!"

assert answer_one().shape == (15,20), "Q1: Your DataFrame should have 20␣
 ↪columns and 15 entries!"
```

```
       ␣
↪---------------------------------------------------------------------------
```

```
    NameError                                 Traceback (most recent call␣
↪last)

    <ipython-input-2-e4f26ba1c257> in <module>
  ----> 1 assert type(answer_one()) == pd.DataFrame, "Q1: You should return a␣
↪DataFrame!"
      2
      3 assert answer_one().shape == (15,20), "Q1: Your DataFrame should␣
↪have 20 columns and 15 entries!"


    <ipython-input-1-bab2565315dc> in answer_one()
      1 def answer_one():
      2     # YOUR CODE HERE
  ----> 3     df=pd.read_excel('assets/Energy Indicators.
↪xls',header=1,skipfooter=1)
      4     df=df.drop(['Unnamed: 0','Unnamed: 1'],axis=1)
      5     for i in range(243,280):


    NameError: name 'pd' is not defined
```

[ ]:
```
# Cell for autograder.
```

1.0.2 Question 2

The previous question joined three datasets then reduced this to just the top 15 entries. When you joined the datasets, but before you reduced this to the top 15 items, how many entries did you lose?

This function should return a single number.

[ ]:
```
%%HTML
<svg width="800" height="300">
  <circle cx="150" cy="180" r="80" fill-opacity="0.2" stroke="black"␣
↪stroke-width="2" fill="blue" />
  <circle cx="200" cy="100" r="80" fill-opacity="0.2" stroke="black"␣
↪stroke-width="2" fill="red" />
  <circle cx="100" cy="100" r="80" fill-opacity="0.2" stroke="black"␣
↪stroke-width="2" fill="green" />
  <line x1="150" y1="125" x2="300" y2="150" stroke="black" stroke-width="2"␣
↪fill="black" stroke-dasharray="5,3"/>
  <text x="300" y="165" font-family="Verdana" font-size="35">Everything but␣
↪this!</text>
</svg>
```

[25]:
```python
def answer_two():
    return int(176)
answer_two()
```

[25]: 176

[ ]:
```python
assert type(answer_two()) == int, "Q2: You should return an int number!"
```

1.0.3 Question 3

What are the top 15 countries for average GDP over the last 10 years?

This function should return a Series named `avgGDP` with 15 countries and their average GDP sorted in descending order.

[38]:
```python
def answer_three():
    GDP=answer_one()
    
↪GDPmean=GDP[['2006','2007','2008','2009','2010','2011','2012','2013','2014','2015']]
    def meangdp(row):
        
↪cols=['2006','2007','2008','2009','2010','2011','2012','2013','2014','2015']
        data=row[cols]
        return pd.Series({'mean': np.mean(data)})
    GDPmean=GDPmean.apply(meangdp,axis=1)
    GDPmean=GDPmean.sort_values(ascending=False,by='mean')
    return GDPmean['mean']
answer_three()
```

[38]:
```
Country
United States         1.536434e+13
China                 6.348609e+12
Japan                 5.542208e+12
Germany               3.493025e+12
France                2.681725e+12
United Kingdom        2.487907e+12
Brazil                2.189794e+12
Italy                 2.120175e+12
India                 1.769297e+12
Canada                1.660647e+12
Russian Federation    1.565459e+12
Spain                 1.418078e+12
Australia             1.164043e+12
South Korea           1.106715e+12
Iran                  4.441558e+11
Name: mean, dtype: float64
```

[ ]:
```python
assert type(answer_three()) == pd.Series, "Q3: You should return a Series!"
```

1.0.4 Question 4

By how much had the GDP changed over the 10 year span for the country with the 6th largest average GDP?

This function should return a single number.

[46]:
```
def answer_four():
    # YOUR CODE HERE
    GDP=answer_one()
    numerateur=GDP.loc['United Kingdom']['2015']-GDP.loc['United␣
↪Kingdom']['2006']
    #denominateur=GDP.loc['United Kingdom']['2006']
    return numerateur
answer_four()
```

[46]:
```
246702696075.3999
```

[47]:
```
# Cell for autograder.
```

1.0.5 Question 5

What is the mean energy supply per capita?

This function should return a single number.

[32]:
```
def answer_five():
    # YOUR CODE HERE
    meanEnergy=answer_one()
    mean=np.mean(meanEnergy['Energy Supply per Capita'])
    return mean
answer_five()
```

[32]:
```
157.6
```

[ ]:
```
# Cell for autograder.
```

1.0.6 Question 6

What country has the maximum % Renewable and what is the percentage?

This function should return a tuple with the name of the country and the percentage.

[33]:
```
def answer_six():
    import numpy as np# YOUR CODE HERE
    maxrenew=answer_one()
    renew=maxrenew['% Renewable']
    renew=renew.astype('float64')
    maxi=np.max(maxrenew['% Renewable'])
    idx=renew.idxmax(skipna=True)
    return (idx,maxi)
answer_six()
```

[33]:
```
('Brazil', 69.64803)
```

[66]:
```
assert type(answer_six()) == tuple, "Q6: You should return a tuple!"

assert type(answer_six()[0]) == str, "Q6: The first element in your result␣
↪should be the name of the country!"
```

```
69.64803


        ␣
↪---------------------------------------------------------------------------

    NotImplementedError                       Traceback (most recent call␣
↪last)

    <ipython-input-66-daa95b3480e4> in <module>
  ----> 1 assert type(answer_six()) == tuple, "Q6: You should return a tuple!"
        2 
        3 assert type(answer_six()[0]) == str, "Q6: The first element in your␣
↪result should be the name of the country!"


    <ipython-input-65-1dd3f65ccb19> in answer_six()
        4     answer=np.max(maxrenew['% Renewable'])
        5     print(answer)
  ----> 6     raise NotImplementedError()
        7 answer_six()


    NotImplementedError:
```

1.0.7 Question 7

Create a new column that is the ratio of Self-Citations to Total Citations. What is the maximum value for this new column, and what country has the highest ratio?

This function should return a tuple with the name of the country and the ratio.

[34]:
```
def answer_seven():
    # YOUR CODE HERE
    import numpy as np
    ratio=answer_one()
    ratio['ratio']=ratio['Self-citations']/ratio['Citations']
    maxi=np.max(ratio['ratio'])
    cit=ratio['ratio']
    country=cit.idxmax()
    maxi=np.max(ratio['ratio'])
    return (country,maxi)
answer_seven()
```

[34]: ('China', 0.6893126179389422)

[13]:
```
assert type(answer_seven()) == tuple, "Q7: You should return a tuple!"

assert type(answer_seven()[0]) == str, "Q7: The first element in your result␣
↪should be the name of the country!"
```

```
      ␣
↪---------------------------------------------------------------------------

        NotImplementedError                       Traceback (most recent call␣
↪last)

        <ipython-input-13-a29f0eae5d14> in <module>
    ----> 1 assert type(answer_seven()) == tuple, "Q7: You should return a tuple!
↪"
          2
          3 assert type(answer_seven()[0]) == str, "Q7: The first element in␣
↪your result should be the name of the country!"


        <ipython-input-12-f5f91fb135c5> in answer_seven()
          7     cit=ratio['ratio']
          8     country=cit.idxmax()
    ----> 9     raise NotImplementedError()
         10     return country
         11 answer_seven()


        NotImplementedError:
```

1.0.8 Question 8

Create a column that estimates the population using Energy Supply and Energy Supply per capita. What is the third most populous country according to this estimate?

This function should return the name of the country

[35]:
```
def answer_eight():
    # YOUR CODE HERE
    pop=answer_one()
    pop=pop.reset_index()
    pop['population']=pop['Energy Supply']/pop['Energy Supply per Capita']
    pop=pop.sort_values(by='population',ascending=False)
    return pop.iloc[2]['Country']
answer_eight()
```

[35]: 'United States'

[ ]:
```
assert type(answer_eight()) == str, "Q8: You should return the name of the␣
↪country!"
```

1.0.9 Question 9

Create a column that estimates the number of citable documents per person. What is the correlation between the number of citable documents per capita and the energy supply per capita? Use the `.corr()` method, (Pearson's correlation).

This function should return a single number.

(Optional: Use the built-in function `plot9()` to visualize the relationship between Energy Supply per Capita vs. Citable docs per Capita)

[42]:
```
def answer_nine():
    # YOUR CODE HERE
    import scipy.stats as stats
    Top15=answer_one()
    Top15['PopEst'] = Top15['Energy Supply'] / Top15['Energy Supply per Capita']
    Top15['Citable docs per Capita'] = Top15['Citable documents'] /␣
↪Top15['PopEst']
    CORR, PVAL=stats.pearsonr(Top15['Citable docs per Capita'],Top15['PopEst'])
    def plot9():
        import matplotlib as plt
        %matplotlib inline

        Top15 = answer_one()
        Top15['PopEst'] = Top15['Energy Supply'] / Top15['Energy Supply per␣
↪Capita']
        Top15['Citable docs per Capita'] = Top15['Citable documents'] /␣
↪Top15['PopEst']
        Top15['Citable docs per Capita']=Top15['Citable docs per Capita'].
↪replace(np.nan,0)
        Top15.plot(x='Citable docs per Capita', y='Energy Supply per Capita',␣
↪kind='scatter', xlim=[0, 0.0006])
    return CORR
answer_nine()
```

[42]: -0.5296539178693623

[89]:
```
def plot9():
    import matplotlib as plt
    %matplotlib inline

    Top15 = answer_one()
    Top15['PopEst'] = Top15['Energy Supply'] / Top15['Energy Supply per Capita']
    Top15['Citable docs per Capita'] = Top15['Citable documents'] /␣
↪Top15['PopEst']
    Top15.plot(x='Citable docs per Capita', y='Energy Supply per Capita',␣
↪kind='scatter', xlim=[0, 0.0006])
```

[ ]:
```
assert answer_nine() >= -1. and answer_nine() <= 1., "Q9: A valid correlation␣
↪should between -1 to 1!"
```

1.0.10 Question 10

Create a new column with a 1 if the country's % Renewable value is at or above the median for all countries in the top 15, and a 0 if the country's % Renewable value is below the median.

This function should return a series named `HighRenew` *whose index is the country name sorted in ascending order of rank.*

[11]:
```
def answer_ten():
    # YOUR CODE HERE
    Top15=answer_one()
    Top15=Top15.reset_index()
    med=Top15['% Renewable'].median(skipna=True)
    def fonction(row,m=med):
        data=row['% Renewable']
        if data>=m:
            row['HighRenew']=1
        else:
            row['HighRenew']=0
        return row
    serie=Top15.apply(fonction,axis=1)
    serie=serie.set_index('Country')
    return serie['HighRenew']
answer_ten()
```

[11]:
```
Country
China                 1
United States         0
Japan                 0
United Kingdom        0
Russian Federation    0
Canada                1
Germany               1
India                 0
France                0
South Korea           0
Italy                 1
Spain                 1
Iran                  0
Australia             0
Brazil                1
Name: HighRenew, dtype: int64
```

[ ]:
```
assert type(answer_ten()) == pd.Series, "Q10: You should return a Series!"
```

1.0.11 Question 11

Use the following dictionary to group the Countries by Continent, then create a DataFrame that displays the sample size (the number of countries in each continent bin), and the sum, mean, and std deviation for the estimated population of each country.

```
ContinentDict  = {'China':'Asia',
                  'United States':'North America',
                  'Japan':'Asia',
                  'United Kingdom':'Europe',
                  'Russian Federation':'Europe',
                  'Canada':'North America',
                  'Germany':'Europe',
                  'India':'Asia',
                  'France':'Europe',
                  'South Korea':'Asia',
                  'Italy':'Europe',
                  'Spain':'Europe',
                  'Iran':'Asia',
                  'Australia':'Australia',
                  'Brazil':'South America'}
```

This function should return a DataFrame with index named Continent `['Asia', 'Australia', 'Europe', 'North America', 'South America']` and columns `['size', 'sum', 'mean', 'std']`

[41]:
```
def answer_eleven():
    # YOUR CODE HERE
    Top15=answer_one()
    ContinentDict  = {'China':'Asia',
                      'United States':'North America',
                      'Japan':'Asia',
                      'United Kingdom':'Europe',
                      'Russian Federation':'Europe',
                      'Canada':'North America',
                      'Germany':'Europe',
                      'India':'Asia',
                      'France':'Europe',
                      'South Korea':'Asia',
                      'Italy':'Europe',
                      'Spain':'Europe',
                      'Iran':'Asia',
                      'Australia':'Australia',
                      'Brazil':'South America'}
    Top15['PopEst'] = Top15['Energy Supply'] / Top15['Energy Supply per Capita']
    Top15['PopEst']=Top15['PopEst'].astype('float64')
    #print(Top15['PopEst'])
    size=Top15.groupby(by=ContinentDict).agg({'PopEst':('count',np.sum,np.
 ↪mean,np.std)})
```

```
    size.columns=['size','sum','mean','std']
    size.index.names = ['Continent']
    return size
answer_eleven()
```

[41]:

```
               size           sum          mean           std
Continent
Asia              4  2.821591e+09  7.053977e+08  7.138779e+08
Australia         1  2.331602e+07  2.331602e+07           NaN
Europe            5  3.940587e+08  7.881174e+07  3.813228e+07
North America     1  3.523986e+07  3.523986e+07           NaN
South America     1  2.059153e+08  2.059153e+08           NaN
```

[ ]:

```
assert type(answer_eleven()) == pd.DataFrame, "Q11: You should return a␣
 ↪DataFrame!"

assert answer_eleven().shape[0] == 5, "Q11: Wrong row numbers!"

assert answer_eleven().shape[1] == 4, "Q11: Wrong column numbers!"
```

1.0.12 Question 12

Cut % Renewable into 5 bins. Group Top15 by the Continent, as well as these new % Renewable bins. How many countries are in each of these groups?

This function should return a Series with a MultiIndex of `Continent`, then the bins for `% Renewable`. Do not include groups with no countries.

[48]:

```
def answer_twelve():
    # YOUR CODE HERE
    Top15=answer_one()
    Top15=Top15.reset_index()
    ContinentDict  = {'China':'Asia',
                  'United States':'North America',
                  'Japan':'Asia',
                  'United Kingdom':'Europe',
                  'Russian Federation':'Europe',
                  'Canada':'North America',
                  'Germany':'Europe',
                  'India':'Asia',
                  'France':'Europe',
                  'South Korea':'Asia',
                  'Italy':'Europe',
                  'Spain':'Europe',
                  'Iran':'Asia',
                  'Australia':'Australia',
                  'Brazil':'South America'}
    result=pd.cut(Top15['% Renewable'],5)
    result.name='Renewable bins'
```

```
    result2=pd.merge(Top15,result,left_index=True,right_index=True)
    result2['Renewable bin']=result2['Renewable bins']
    #result2=result2.set_index(['Country'])
    result2['Continent']=result2['Country'].map(ContinentDict)
    answer=result2.groupby(by=['Continent', 'Renewable bins']).agg({'Renewable␣
↪bin':'count'})
    answer=answer.rename_axis(index=['Continent', '% Renewable'])
    return answer['Renewable bin']
answer_twelve()
```

[48]:
```
Continent      % Renewable
Asia           (2.212, 15.753]     4
               (15.753, 29.227]    1
Australia      (2.212, 15.753]     1
Europe         (2.212, 15.753]     1
               (15.753, 29.227]    3
               (29.227, 42.701]    2
North America  (2.212, 15.753]     1
               (56.174, 69.648]    1
South America  (56.174, 69.648]    1
Name: Renewable bin, dtype: int64
```

[ ]:
```
assert type(answer_twelve()) == pd.Series, "Q12: You should return a Series!"

assert len(answer_twelve()) == 9, "Q12: Wrong result numbers!"
```

1.0.13 Question 13

Convert the Population Estimate series to a string with thousands separator (using commas). Use all significant digits (do not round the results).

e.g. 12345678.90 -> 12,345,678.90

This function should return a series `PopEst` *whose index is the country name and whose values are the population estimate string*

[43]:
```
def answer_thirteen():
    # YOUR CODE HERE
    Top15=answer_one()
    Top15['PopEst'] = Top15['Energy Supply'] / Top15['Energy Supply per Capita']
    Top15['PopEst']=Top15['PopEst'].astype('float64')
    Top15['PopEst']=Top15['PopEst'].apply('{:,}'.format)
    return Top15['PopEst']
    raise NotImplementedError()
answer_thirteen()
```

[43]:
```
Country
China                 1,367,645,161.2903225
United States                           nan
Japan                   127,409,395.97315437
United Kingdom                          nan
```

```
Russian Federation             143,500,000.0
Canada                       35,239,864.86486486
Germany                      80,369,696.96969697
India                      1,276,730,769.2307692
France                       63,837,349.39759036
South Korea                 49,805,429.864253394
Italy                       59,908,256.880733944
Spain                         46,443,396.2264151
Iran                                         nan
Australia                   23,316,017.316017315
Brazil                      205,915,254.23728815
Name: PopEst, dtype: object
```

[ ]:
```
assert type(answer_thirteen()) == pd.Series, "Q13: You should return a Series!"

assert len(answer_thirteen()) == 15, "Q13: Wrong result numbers!"
```

1.0.14 Optional

Use the built in function `plot_optional()` to see an example visualization.

[ ]:
```
def plot_optional():
    import matplotlib as plt
    %matplotlib inline
    Top15 = answer_one()
    ax = Top15.plot(x='Rank', y='% Renewable', kind='scatter',
                    ␣
 ↪c=['#e41a1c','#377eb8','#e41a1c','#4daf4a','#4daf4a','#377eb8','#4daf4a','#e41a1c',
                       ␣
 ↪'#4daf4a','#e41a1c','#4daf4a','#4daf4a','#e41a1c','#dede00','#ff7f00'],
                    xticks=range(1,16), s=6*Top15['2014']/10**10, alpha=.75,␣
 ↪figsize=[16,6]);

    for i, txt in enumerate(Top15.index):
        ax.annotate(txt, [Top15['Rank'][i], Top15['% Renewable'][i]],␣
 ↪ha='center')

    print("This is an example of a visualization that can be created to help␣
 ↪understand the data. \
This is a bubble chart showing % Renewable vs. Rank. The size of the bubble␣
 ↪corresponds to the countries' \
2014 GDP, and the color corresponds to the continent.")
```