Program Summary - Project 3.sas

Execution Environment

Author:	bdomash@wisc.edu
File:	/home/bdomash0/my_courses/Homeworks/Project 3/Project 3.sas
SAS Platform:	Linux LIN X64 3.10.0-693.21.1.el7.x86_64
SAS Host:	ODAWS01.ODA.SAS.COM
SAS Version:	9.04.01M5P09132017
SAS Locale:	en_US
Submission Time:	5/9/2019, 7:20:15 PM
Browser Host:	63.163.122.27
User Agent:	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_0) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/73.0.3683.103 Safari/537.36
Application Server:	ODAMID00-PROD-US.ODA.SAS.COM

Code: Project 3.sas

```
*Part 1: Read in the three datasets;

*Dataset 1: containing State names and Obesity Rates;
%let path = https://raw.githubusercontent.com/bdomash/SAS_project_data/master/;

filename obesity
    url "&path.Obesity_by_state.csv"
    termstr=lf ;

data obesity;
    infile obesity dlm=',' firstobs=1 dsd missover;
    input state :$25. rate;
run;
*dataset 2: containing Location info for each McDonalds in the US;

filename mcdons
    url "&path.mcdonalds.csv"
    termstr=lf;

data mcdonalds;
    infile mcdons dlm=',' firstobs=1 dsd missover;
    input longitude lattitude name :$199. address :$199.;
run;

*dataset 3: list of states with their abbreviations to later convert within the two above datasets;
filename abbrev
    url "&path.states.csv"
    termstr=lf;

data abbreviations;
    infile abbrev dlm = ',' firstobs=2 dsd missover;
    input long :$27. short :$2.;
run;

*dataset 4: states and populations;
filename popu
    url "&path.population.csv"
    termstr=lf;

data population;
    infile popu dlm = ',' firstobs=2 dsd missover;
    input state :$27. population;
run;


*Part 2: Data-wrangling;



*First clean up mcdonalds dataset, pull out city and state information
For some observations, the state data pulls out erroneous info.
Taking first 2 letters takes just the state abbreviation;
data mcdonalds;
    set mcdonalds;
    city = scan(address,2,",");
    state = scan(address,3,",");
    state = substrn(state,1,2);
run;
```

```
*Sort each of the datasets by full state name for merging;
proc sort data=abbreviations;by long;run;
proc sort data=obesity;by state;run;
proc sort data=population;by state;run;

*Adding abbreviated state to obesity dataset for later merge with with mcdonalds data;
*We will not need Guam or PR for these analyses;
data obesity;
    merge obesity(in=a) abbreviations(rename=(long=state) in=b);
    by state;
    if a = 1;
    if state = 'Guam' then delete;
    if state = 'Puerto Rico' then delete;
    rename short = Abbreviation;
    run;

*Add abbrevaiated state to population for later merge with mcdonalds data;
*Remove erroneous observations such as United States and Regions that were present;
data population;
    merge population(in=a) abbreviations(rename=(long=state));
    by state;
    if a = 1;
    if short='' then delete;
    rename state=long;
    rename short = state;
    run;


*Part 3: analysis;



*First, let's see which cities have the most McDonalds and map them;
*Mcdonalds dataset seperates NYC into 5 buroughs, instead lets include them all as NYC;
data temp;
    set mcdonalds;
    if findw(address,'Brooklyn,NY')>0 then city = 'New York';
    if findw(address,'Queens,NY')>0 then city = 'New York';
    if findw(address,'Staten Island,NY')>0 then city = 'New York';
    if findw(address,'Manhattan,NY')>0 then city = 'New York';
    if findw(address,'Bronx,NY')>0 then city = 'New York';
    if findw(address,'New York,NY')>0 then city = 'New York';
run;

*Grouping data, counting how many mcdonalds locations for each city, state pair;
proc sql noprint;
    create table cities as
    select city, state, count(1) as count from temp
    group by city, state;
quit;

*Sorting two datasets for merge;
proc sort data=cities;by state city;run;
proc sort data=mcdonalds; by state city;run;

*Now, for each city, state pair we are going to add a location from the mcdonalds dataset
While the coordinates will not be exact, it will give us a general location for mapping.
We only want one coordinate location for each city, state pair
And finally we need to manually add in NY's location as it was missing;
data temp;
    merge cities(in=a) mcdonalds(in=b);
    by state city;
    if a=1;
    if first.city=1;
    drop name address;
    if city = 'New York' then do;
        lattitude = 40.758805;
        longitude = -73.984727;
        end;
    lab = cats(city,' (',count,')');
    run;


*Sorting the data to display the cities with the most Mcdonalds locations first;
proc sort data=temp;by descending count;run;


*Mapping the 10 cities with the most mcdonalds;
PROC SGMAP plotdata=temp(obs=10);
    openstreetmap;
    TITLE H=2 "Cities with the Most McDonalds";
```

```
    scatter X=longitude Y=lattitude / MARKERATTRS=(COLOR=cxff3344  symbol = CircleFilled SIZE= 10)
        datalabel = lab DATALABELATTRS=(COLOR=cxff3344 Weight=Bold SIZE=10) DATALABELPOS=Left;
RUN;

proc print data=temp(obs=10);
    var city state count;
run;

*Next, lets look at the states with the most McDonalds per capita.
Here, we group the Mcdonalds locations by states, counting how many locations per state;
proc sql noprint;
    create table states as
    select state, count(1) as count from mcdonalds
    group by state;
quit;

*Sorting by state to merge;
proc sort data=population;by state;run;
proc sort data=states;by state; run;

*Here, we merge our state-grouped McDonalds data with each state's population
Now we have Mcdonalds per state and population per state in each row
We also create a per-capita variable, using 100000 since it creates nice single-digit values;
data states;
    merge states(in=a) population(in=b);
    by state;
    if a=1 and b=1;
    per_100000 = count/population*100000;
    run;

*Sorting the dataset by states with most McDonalds per capita;
proc sort data=states; by descending per_100000;run;


*Plotting a barchart of the states with the most McDonalds per-capita;
title "States with the most McDonalds per 100,000 people";
        proc sgplot data = states(obs=10);
            xaxis label = 'State';
            yaxis label = 'McDonalds per 100,000';
            vbar state / datalabel response=per_100000 CATEGORYORDER=RESPDESC
            datalabelattrs=(size=12pt)
            fillattrs=(color='blue');
        run;
*Lets print the results as well to better visualize which states have the most McDonalds;
proc print data=states labels;
    var long per_100000;
    label long = 'State' per_100000 = 'McDonalds/100,000';
    run;



*Next lets plot the states with the highest obesity rates;
proc sort data=obesity;by descending rate;run;
title "States with the Highest Obesity Rate";
proc sgplot data = obesity(obs=10);
            xaxis label = 'State';
            yaxis label = 'Obesity Rate';
            vbar Abbreviation / datalabel response=rate CATEGORYORDER=RESPDESC
            datalabelattrs=(size=12pt)
            fillattrs=(color='red');
        run;
*Lets also print these results;
proc print data=obesity label;
    var state rate;
    label state = 'State' rate = 'Obesity Rate (%)';
    run;

*Finally lets look at the relationship between the two;
*First we must merge the obesity state-wide data with the McDonalds state-wide per capita data;
proc sort data=obesity;by state;run;
proc sort data=states;by long;run;
data merged;
    merge obesity(rename=(state=long)) states(in=b);
    by long;
    drop abbreviation;
    run;


*Now we create a scatter plot between the two variables. We include a regression line;
title 'Obesity Rate vs McDonalds per 100000 people';
proc sgplot data=merged;
```

```
    xaxis label = 'McDonalds per 100000 people';
    yaxis label = 'Obesity Rate';
    reg x=per_100000 y=rate / lineattrs=(color=red thickness=2) datalabel=state;
    run;

*This visualization would be a lot better if we could color code by region;
*Lets add some more data;
filename region
    url "https://raw.githubusercontent.com/cphalpert/census-regions/master/us%20census%20bureau%20regions%20and%20d
    termstr=lf ;

data region;
    infile region dlm=',' firstobs=2 dsd missover;
    input state :$25. short :$2. region :$10. division :$25.;
run;

proc sort data=region;by state;run;
proc sort data=merged;by long; run;
data merged;
    merge merged(in=a) region(in=b rename=(state=long));
    by long;
    drop short;
run;

*Plot again with region;
title 'Obesity Rate vs McDonalds per 100000 people';
proc sgplot data=merged;
    styleattrs datacontrastcolors=(red green orange blue);
    xaxis label = 'McDonalds per 100000 people';
    yaxis label = 'Obesity Rate';
    scatter x=per_100000 y=rate / group=region markerattrs=(symbol=CircleFilled) markeroutlineattrs=(color=black th
    reg x=per_100000 y=rate / lineattrs=(color=red thickness=2) datalabel=state;
    run;

*Finally we do some simple regression analysis to check the relationship between the two variables;
proc reg data=merged;
    label rate ='Obesity Rate';
    label per_100000 ='McDonalds per 1000000';
    model rate=per_100000;
run;

proc corr data=merged NOMISS plots=matrix;
    var rate per_100000;
run;

*The analysis shows that there is a significant relationship between the two variables;
*However, there appear to be (at least) two outlier variables
Lets remove those and see if this affects the analysis;
data no_outliers;
    set merged;
    if state = 'DC' then delete;
    if state = 'HI' then delete;
    *if state = 'CO' then delete;
    *if state = 'MT' then delete;
    *if state = 'NV' then delete;
run;

*Once again we can see a significant relationship between the two
We can conclude that there is a significant relationship between McDonalds in a state and a state's obesity rate;
proc reg data=no_outliers;
    label rate ='Obesity Rate';
    label per_100000 ='McDonalds per 1000000';
    model rate=per_100000;
run;

proc corr data=no_outliers NOMISS plots=matrix;
    var rate per_100000;
run;
```

Log: Project 3.sas

Warnings (5)

Notes (118)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
70
71         *Part 1: Read in the three datasets;
72
73         *Dataset 1: containing State names and Obesity Rates;
```

```
74         %let path = https://raw.githubusercontent.com/bdomash/SAS_project_data/master/;
75
76         filename obesity
77         url "&path.Obesity_by_state.csv"
78         termstr=lf ;
79
80         data obesity;
81         infile obesity dlm=',' firstobs=1 dsd missover;
82         input state :$25. rate;
83         run;

NOTE: The infile OBESITY is:
      Filename=https://raw.githubusercontent.com/bdomash/SAS_project_data/master/Obesity_by_state.csv,
      Local Host Name=odaws02-prod-us,
      Local Host IP addr=10.249.126.103,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.52.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: 53 records were read from the infile OBESITY.
      The minimum record length was 8.
      The maximum record length was 25.
NOTE: The data set WORK.OBESITY has 53 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.32 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              523.75k
      OS Memory           39852.00k
      Timestamp           05/10/2019 12:19:41 AM
      Step Count                        130  Switch Count  6
      Page Faults                       0
      Page Reclaims                     101
      Page Swaps                        0
      Voluntary Context Switches        36
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


84         *dataset 2: containing Location info for each McDonalds in the US;
85
86         filename mcdons
87         url "&path.mcdonalds.csv"
88         termstr=lf;
89
90         data mcdonalds;
91         infile mcdons dlm=',' firstobs=1 dsd missover;
92         input longitude lattitude name :$199. address :$199.;
93         run;

NOTE: The infile MCDONS is:
      Filename=https://raw.githubusercontent.com/bdomash/SAS_project_data/master/mcdonalds.csv,
      Local Host Name=odaws02-prod-us,
      Local Host IP addr=10.249.126.103,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.52.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: 13912 records were read from the infile MCDONS.
      The minimum record length was 65.
      The maximum record length was 199.
NOTE: The data set WORK.MCDONALDS has 13912 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.27 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              1796.31k
      OS Memory           40364.00k
      Timestamp           05/10/2019 12:19:41 AM
      Step Count                        131  Switch Count  14
      Page Faults                       0
      Page Reclaims                     134
      Page Swaps                        0
      Voluntary Context Switches        83
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           11528


94
95         *dataset 3: list of states with their abbreviations to later convert within the two above datasets;
96         filename abbrev
97         url "&path.states.csv"
98         termstr=lf;
99
100        data abbreviations;
101        infile abbrev dlm = ',' firstobs=2 dsd missover;
102        input long :$27. short :$2.;
103        run;
```

```
NOTE: The infile ABBREV is:
      Filename=https://raw.githubusercontent.com/bdomash/SAS_project_data/master/states.csv,
      Local Host Name=odaws02-prod-us,
      Local Host IP addr=10.249.126.103,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.52.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: 51 records were read from the infile ABBREV.
      The minimum record length was 11.
      The maximum record length was 27.
NOTE: The data set WORK.ABBREVIATIONS has 51 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.19 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              523.75k
      OS Memory           39852.00k
      Timestamp           05/10/2019 12:19:41 AM
      Step Count                        132  Switch Count  6
      Page Faults                       0
      Page Reclaims                     95
      Page Swaps                        0
      Voluntary Context Switches        35
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


104
105        *dataset 4: states and populations;
106        filename popu
107        url "&path.population.csv"
108        termstr=lf;
109
110        data population;
111        infile popu dlm = ',' firstobs=2 dsd missover;
112        input state :$27. population;
113        run;

NOTE: The infile POPU is:
      Filename=https://raw.githubusercontent.com/bdomash/SAS_project_data/master/population.csv,
      Local Host Name=odaws02-prod-us,
      Local Host IP addr=10.249.126.103,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.52.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: 57 records were read from the infile POPU.
      The minimum record length was 12.
      The maximum record length was 27.
NOTE: The data set WORK.POPULATION has 57 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.18 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              523.75k
      OS Memory           39852.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        133  Switch Count  6
      Page Faults                       0
      Page Reclaims                     93
      Page Swaps                        0
      Voluntary Context Switches        35
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


114
115
116        *Part 2: Data-wrangling;
117
118
119
120        *First clean up mcdonalds dataset, pull out city and state information
121        For some observations, the state data pulls out erroneous info.
122        Taking first 2 letters takes just the state abbreviation;
123        data mcdonalds;
124        set mcdonalds;
125        city = scan(address,2,",");
126        state = scan(address,3,",");
127        state = substrn(state,1,2);
128        run;

NOTE: There were 13912 observations read from the data set WORK.MCDONALDS.
NOTE: The data set WORK.MCDONALDS has 13912 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
```

```
      user cpu time       0.01 seconds
      system cpu time     0.02 seconds
      memory              3416.81k
      OS Memory           42416.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        134  Switch Count  2
      Page Faults                       0
      Page Reclaims                     496
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           22280


129
130        *Sort each of the datasets by full state name for merging;
131        proc sort data=abbreviations;by long;run;

NOTE: There were 51 observations read from the data set WORK.ABBREVIATIONS.
NOTE: The data set WORK.ABBREVIATIONS has 51 observations and 2 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              783.00k
      OS Memory           40112.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        135  Switch Count  2
      Page Faults                       0
      Page Reclaims                     114
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


132        proc sort data=obesity;by state;run;

NOTE: There were 53 observations read from the data set WORK.OBESITY.
NOTE: The data set WORK.OBESITY has 53 observations and 2 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              779.00k
      OS Memory           40112.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        136  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


133        proc sort data=population;by state;run;

NOTE: There were 57 observations read from the data set WORK.POPULATION.
NOTE: The data set WORK.POPULATION has 57 observations and 2 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              779.00k
      OS Memory           40112.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        137  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


134
135        *Adding abbreviated state to obesity dataset for later merge with with mcdonalds data;
136        *We will not need Guam or PR for these analyses;
137        data obesity;
138        merge obesity(in=a) abbreviations(rename=(long=state) in=b);
139        by state;
140        if a = 1;
141        if state = 'Guam' then delete;
142        if state = 'Puerto Rico' then delete;
143        rename short = Abbreviation;
144        run;
```

```
WARNING: Multiple lengths were specified for the BY variable state by input data sets. This might cause unexpected results.
NOTE: There were 53 observations read from the data set WORK.OBESITY.
NOTE: There were 51 observations read from the data set WORK.ABBREVIATIONS.
NOTE: The data set WORK.OBESITY has 51 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1242.65k
      OS Memory           40372.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        138  Switch Count  2
      Page Faults                       0
      Page Reclaims                     162
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


145
146        *Add abbrevaiated state to population for later merge with mcdonalds data;
147        *Remove erroneous observations such as United States and Regions that were present;
148        data population;
149        merge population(in=a) abbreviations(rename=(long=state));
150        by state;
151        if a = 1;
152        if short='' then delete;
153        rename state=long;
154        rename short = state;
155        run;

NOTE: There were 57 observations read from the data set WORK.POPULATION.
NOTE: There were 51 observations read from the data set WORK.ABBREVIATIONS.
NOTE: The data set WORK.POPULATION has 51 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1242.65k
      OS Memory           40372.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        139  Switch Count  2
      Page Faults                       0
      Page Reclaims                     156
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


156
157
158        *Part 3: analysis;
159
160
161
162        *First, let's see which cities have the most McDonalds and map them;
163        *Mcdonalds dataset seperates NYC into 5 buroughs, instead lets include them all as NYC;
164        data temp;
165        set mcdonalds;
166        if findw(address,'Brooklyn,NY')>0 then city = 'New York';
167        if findw(address,'Queens,NY')>0 then city = 'New York';
168        if findw(address,'Staten Island,NY')>0 then city = 'New York';
169        if findw(address,'Manhattan,NY')>0 then city = 'New York';
170        if findw(address,'Bronx,NY')>0 then city = 'New York';
171        if findw(address,'New York,NY')>0 then city = 'New York';
172        run;

NOTE: There were 13912 observations read from the data set WORK.MCDONALDS.
NOTE: The data set WORK.TEMP has 13912 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.04 seconds
      user cpu time       0.04 seconds
      system cpu time     0.01 seconds
      memory              3444.87k
      OS Memory           42416.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        140  Switch Count  2
      Page Faults                       0
      Page Reclaims                     494
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           22280


173
174        *Grouping data, counting how many mcdonalds locations for each city, state pair;
```

```
175        proc sql noprint;
176        create table cities as
177            select city, state, count(1) as count from temp
178            group by city, state;
NOTE: Table WORK.CITIES created, with 6389 rows and 3 columns.

179        quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.02 seconds
      memory              17301.32k
      OS Memory           54728.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        141  Switch Count  3
      Page Faults                       0
      Page Reclaims                     3523
      Page Swaps                        0
      Voluntary Context Switches        160
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           5128


180
181        *Sorting two datasets for merge;
182        proc sort data=cities;by state city;run;

NOTE: There were 6389 observations read from the data set WORK.CITIES.
NOTE: The data set WORK.CITIES has 6389 observations and 3 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              8440.37k
      OS Memory           47056.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        142  Switch Count  2
      Page Faults                       0
      Page Reclaims                     1600
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           5128


183        proc sort data=mcdonalds; by state city;run;

NOTE: There were 13912 observations read from the data set WORK.MCDONALDS.
NOTE: The data set WORK.MCDONALDS has 13912 observations and 6 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.02 seconds
      user cpu time       0.01 seconds
      system cpu time     0.02 seconds
      memory              19893.03k
      OS Memory           58364.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        143  Switch Count  2
      Page Faults                       0
      Page Reclaims                     4494
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           22280


184
185        *Now, for each city, state pair we are going to add a location from the mcdonalds dataset
186        While the coordinates will not be exact, it will give us a general location for mapping.
187        We only want one coordinate location for each city, state pair
188        And finally we need to manually add in NY's location as it was missing;
189        data temp;
190        merge cities(in=a) mcdonalds(in=b);
191        by state city;
192        if a=1;
193        if first.city=1;
194        drop name address;
195        if city = 'New York' then do;
196        lattitude = 40.758805;
197        longitude = -73.984727;
198        end;
199        lab = cats(city,' (',count,')');
200        run;

NOTE: There were 6389 observations read from the data set WORK.CITIES.
NOTE: There were 13912 observations read from the data set WORK.MCDONALDS.
NOTE: The data set WORK.TEMP has 6389 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
```

```
      system cpu time     0.01 seconds
      memory              5222.78k
      OS Memory           44468.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        144  Switch Count  2
      Page Faults                       0
      Page Reclaims                     866
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           7944


201
202
203        *Sorting the data to display the cities with the most Mcdonalds locations first;
204        proc sort data=temp;by descending count;run;

NOTE: There were 6389 observations read from the data set WORK.TEMP.
NOTE: The data set WORK.TEMP has 6389 observations and 6 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              7372.75k
      OS Memory           46028.00k
      Timestamp           05/10/2019 12:19:42 AM
      Step Count                        145  Switch Count  2
      Page Faults                       0
      Page Reclaims                     1324
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           7944


205
206
207        *Mapping the 10 cities with the most mcdonalds;
208        PROC SGMAP plotdata=temp(obs=10);
209        openstreetmap;
210        TITLE H=2 "Cities with the Most McDonalds";
211        scatter X=longitude Y=lattitude / MARKERATTRS=(COLOR=cxff3344  symbol = CircleFilled SIZE= 10)
212        datalabel = lab DATALABELATTRS=(COLOR=cxff3344 Weight=Bold SIZE=10) DATALABELPOS=Left;
213        RUN;

NOTE: PROCEDURE SGMAP used (Total process time):
      real time           30.30 seconds
      user cpu time       0.26 seconds
      system cpu time     0.06 seconds
      memory              35178.23k
      OS Memory           72564.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        146  Switch Count  2
      Page Faults                       0
      Page Reclaims                     29656
      Page Swaps                        0
      Voluntary Context Switches        37
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           2456

WARNING: Some features may not be displayed on the map because of missing location information in the data.
WARNING: Some features may not be displayed on the map because of missing location information in the data.
NOTE: There were 10 observations read from the data set WORK.TEMP.

214
215        proc print data=temp(obs=10);
216        var city state count;
217        run;

NOTE: There were 10 observations read from the data set WORK.TEMP.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.02 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              2727.34k
      OS Memory           50400.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        147  Switch Count  0
      Page Faults                       0
      Page Reclaims                     271
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8


218
```

```
219        *Next, lets look at the states with the most McDonalds per capita.
220        Here, we group the Mcdonalds locations by states, counting how many locations per state;
221        proc sql noprint;
222        create table states as
223            select state, count(1) as count from mcdonalds
224            group by state;
NOTE: Table WORK.STATES created, with 55 rows and 2 columns.

225        quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              6779.53k
      OS Memory           55524.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        148  Switch Count  3
      Page Faults                       0
      Page Reclaims                     315
      Page Swaps                        0
      Voluntary Context Switches        12
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


226
227        *Sorting by state to merge;
228        proc sort data=population;by state;run;

NOTE: There were 51 observations read from the data set WORK.POPULATION.
NOTE: The data set WORK.POPULATION has 51 observations and 3 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              787.75k
      OS Memory           49892.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        149  Switch Count  2
      Page Faults                       0
      Page Reclaims                     115
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


229        proc sort data=states;by state; run;

NOTE: Input data set is already sorted, no sorting done.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              391.43k
      OS Memory           49632.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        150  Switch Count  0
      Page Faults                       0
      Page Reclaims                     49
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


230
231        *Here, we merge our state-grouped McDonalds data with each state's population
232        Now we have Mcdonalds per state and population per state in each row
233        We also create a per-capita variable, using 100000 since it creates nice single-digit values;
234        data states;
235        merge states(in=a) population(in=b);
236        by state;
237        if a=1 and b=1;
238        per_100000 = count/population*100000;
239        run;

NOTE: There were 55 observations read from the data set WORK.STATES.
NOTE: There were 51 observations read from the data set WORK.POPULATION.
NOTE: The data set WORK.STATES has 51 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1251.78k
      OS Memory           50152.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        151  Switch Count  2
      Page Faults                       0
```

```
      Page Reclaims                    165
      Page Swaps                       0
      Voluntary Context Switches       10
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          264


240
241        *Sorting the dataset by states with most McDonalds per capita;
242        proc sort data=states; by descending per_100000;run;

NOTE: There were 51 observations read from the data set WORK.STATES.
NOTE: The data set WORK.STATES has 51 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              774.87k
      OS Memory           49892.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                       152  Switch Count  2
      Page Faults                      0
      Page Reclaims                    110
      Page Swaps                       0
      Voluntary Context Switches       10
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          272


243
244
245        *Plotting a barchart of the states with the most McDonalds per-capita;
246        title "States with the most McDonalds per 100,000 people";
247
247      !   proc sgplot data = states(obs=10);
248        xaxis label = 'State';
249        yaxis label = 'McDonalds per 100,000';
250        vbar state / datalabel response=per_100000 CATEGORYORDER=RESPDESC
251        datalabelattrs=(size=12pt)
252        fillattrs=(color='blue');
253        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.13 seconds
      user cpu time       0.05 seconds
      system cpu time     0.01 seconds
      memory              2725.09k
      OS Memory           51180.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                       153  Switch Count  12
      Page Faults                      0
      Page Reclaims                    764
      Page Swaps                       0
      Voluntary Context Switches       263
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          952

NOTE: There were 10 observations read from the data set WORK.STATES.

254        *Lets print the results as well to better visualize which states have the most McDonalds;
255        proc print data=states labels;
                                    _____
                                    1
WARNING 1-322: Assuming the symbol LABEL was misspelled as labels.
256        var long per_100000;
257        label long = 'State' per_100000 = 'McDonalds/100,000';
258        run;

NOTE: There were 51 observations read from the data set WORK.STATES.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.04 seconds
      user cpu time       0.05 seconds
      system cpu time     0.00 seconds
      memory              475.46k
      OS Memory           50400.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                       154  Switch Count  0
      Page Faults                      0
      Page Reclaims                    60
      Page Swaps                       0
      Voluntary Context Switches       0
      Involuntary Context Switches     0
      Block Input Operations           0
      Block Output Operations          8


259
260
261
```

```
262        *Next lets plot the states with the highest obesity rates;
263        proc sort data=obesity;by descending rate;run;

NOTE: There were 51 observations read from the data set WORK.OBESITY.
NOTE: The data set WORK.OBESITY has 51 observations and 3 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              787.25k
      OS Memory           50356.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        155  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


264        title "States with the Highest Obesity Rate";
265        proc sgplot data = obesity(obs=10);
266        xaxis label = 'State';
267        yaxis label = 'Obesity Rate';
268        vbar Abbreviation / datalabel response=rate CATEGORYORDER=RESPDESC
269        datalabelattrs=(size=12pt)
270        fillattrs=(color='red');
271        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.12 seconds
      user cpu time       0.05 seconds
      system cpu time     0.00 seconds
      memory              3081.82k
      OS Memory           51900.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        156  Switch Count  12
      Page Faults                       0
      Page Reclaims                     715
      Page Swaps                        0
      Voluntary Context Switches        262
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           968

NOTE: There were 10 observations read from the data set WORK.OBESITY.

272        *Lets also print these results;
273        proc print data=obesity label;
274        var state rate;
275        label state = 'State' rate = 'Obesity Rate (%)';
276        run;

NOTE: There were 51 observations read from the data set WORK.OBESITY.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.04 seconds
      user cpu time       0.04 seconds
      system cpu time     0.00 seconds
      memory              469.43k
      OS Memory           50608.00k
      Timestamp           05/10/2019 12:20:12 AM
      Step Count                        157  Switch Count  0
      Page Faults                       0
      Page Reclaims                     60
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8

276      !

277
278        *Finally lets look at the relationship between the two;
279        *First we must merge the obesity state-wide data with the McDonalds state-wide per capita data;
280        proc sort data=obesity;by state;run;

NOTE: There were 51 observations read from the data set WORK.OBESITY.
NOTE: The data set WORK.OBESITY has 51 observations and 3 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              781.78k
      OS Memory           50868.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        158  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
```

```
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


281        proc sort data=states;by long;run;

NOTE: There were 51 observations read from the data set WORK.STATES.
NOTE: The data set WORK.STATES has 51 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              774.25k
      OS Memory           50868.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        159  Switch Count  2
      Page Faults                       0
      Page Reclaims                     110
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


282        data merged;
283        merge obesity(rename=(state=long)) states(in=b);
284        by long;
285        drop abbreviation;
286        run;

WARNING: Multiple lengths were specified for the BY variable long by input data sets. This might cause unexpected results.
NOTE: There were 51 observations read from the data set WORK.OBESITY.
NOTE: There were 51 observations read from the data set WORK.STATES.
NOTE: The data set WORK.MERGED has 51 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1248.25k
      OS Memory           51128.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        160  Switch Count  2
      Page Faults                       0
      Page Reclaims                     160
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


287
288
289        *Now we create a scatter plot between the two variables. We include a regression line;
290        title 'Obesity Rate vs McDonalds per 100000 people';
291        proc sgplot data=merged;
292        xaxis label = 'McDonalds per 100000 people';
293        yaxis label = 'Obesity Rate';
294            reg x=per_100000 y=rate / lineattrs=(color=red thickness=2) datalabel=state;
295            run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.15 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              2607.50k
      OS Memory           51512.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        161  Switch Count  1
      Page Faults                       0
      Page Reclaims                     337
      Page Swaps                        0
      Voluntary Context Switches        382
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           568

NOTE: There were 51 observations read from the data set WORK.MERGED.

296
297        *This visualization would be a lot better if we could color code by region;
298        *Lets add some more data;
299        filename region
300        url
300      ! "https://raw.githubusercontent.com/cphalpert/census-regions/master/us%20census%20bureau%20regions%20and%20divisions.csv'
301        termstr=lf ;
302
303        data region;
304        infile region dlm=',' firstobs=2 dsd missover;
```

```
305        input state :$25. short :$2. region :$10. division :$25.;
306        run;

NOTE: The infile REGION is:

      Filename=https://raw.githubusercontent.com/cphalpert/census-regions/master/us%20census%20bureau%20regions%20and%20divisions.c
      v,
      Local Host Name=odaws02-prod-us,
      Local Host IP addr=10.249.126.103,
      Service Hostname Name=raw.githubusercontent.com,
      Service IP addr=151.101.40.133,
      Service Name=N/A,Service Portno=443,
      Lrecl=32767,Recfm=Variable

NOTE: 51 records were read from the infile REGION.
      The minimum record length was 21.
      The maximum record length was 44.
NOTE: The data set WORK.REGION has 51 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.26 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              523.90k
      OS Memory           50608.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        162  Switch Count  6
      Page Faults                       0
      Page Reclaims                     94
      Page Swaps                        0
      Voluntary Context Switches        36
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           264


307
308        proc sort data=region;by state;run;

NOTE: There were 51 observations read from the data set WORK.REGION.
NOTE: The data set WORK.REGION has 51 observations and 4 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              780.18k
      OS Memory           50868.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        163  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


309        proc sort data=merged;by long; run;

NOTE: There were 51 observations read from the data set WORK.MERGED.
NOTE: The data set WORK.MERGED has 51 observations and 6 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              774.28k
      OS Memory           50868.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        164  Switch Count  2
      Page Faults                       0
      Page Reclaims                     110
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


310        data merged;
311        merge merged(in=a) region(in=b rename=(state=long));
312        by long;
313        drop short;
314        run;

NOTE: There were 51 observations read from the data set WORK.MERGED.
NOTE: There were 51 observations read from the data set WORK.REGION.
NOTE: The data set WORK.MERGED has 51 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1251.15k
```

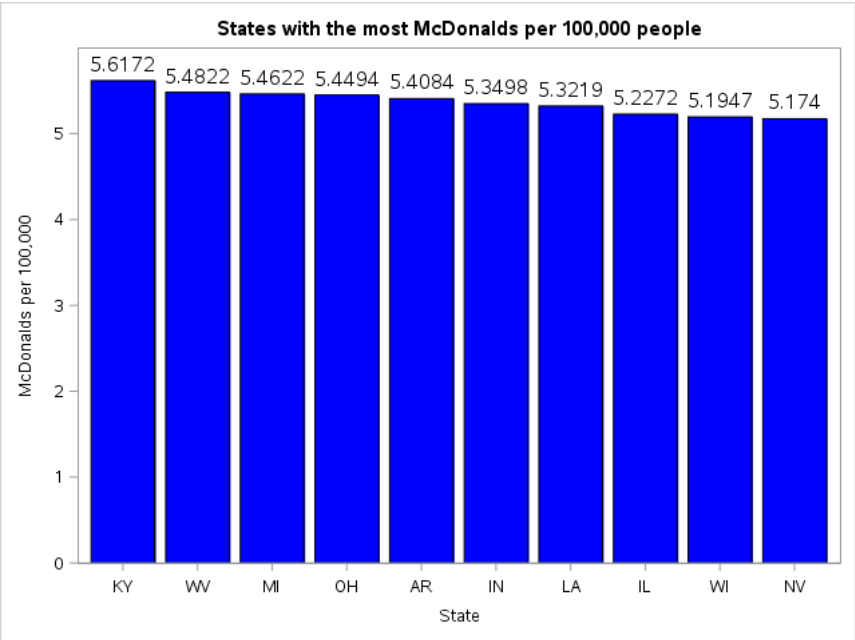
```
      OS Memory           51128.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        165  Switch Count  2
      Page Faults                       0
      Page Reclaims                     157
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


315
316        *Plot again with region;
317        title 'Obesity Rate vs McDonalds per 100000 people';
318        proc sgplot data=merged;
319        styleattrs datacontrastcolors=(red green orange blue);
320        xaxis label = 'McDonalds per 100000 people';
321        yaxis label = 'Obesity Rate';
322        scatter x=per_100000 y=rate / group=region markerattrs=(symbol=CircleFilled) markeroutlineattrs=(color=black
322      ! thickness=1);
323             reg x=per_100000 y=rate / lineattrs=(color=red thickness=2) datalabel=state;
324             run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.18 seconds
      user cpu time       0.06 seconds
      system cpu time     0.01 seconds
      memory              2597.21k
      OS Memory           51768.00k
      Timestamp           05/10/2019 12:20:13 AM
      Step Count                        166  Switch Count  1
      Page Faults                       0
      Page Reclaims                     369
      Page Swaps                        0
      Voluntary Context Switches        547
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           576

NOTE: There were 51 observations read from the data set WORK.MERGED.

325
326        *Finally we do some simple regression analysis to check the relationship between the two variables;
327        proc reg data=merged;
328        label rate ='Obesity Rate';
329        label per_100000 ='McDonalds per 1000000';
330        model rate=per_100000;
331        run;

332

NOTE: PROCEDURE REG used (Total process time):
      real time           0.74 seconds
      user cpu time       0.22 seconds
      system cpu time     0.04 seconds
      memory              12122.50k
      OS Memory           60140.00k
      Timestamp           05/10/2019 12:20:14 AM
      Step Count                        167  Switch Count  22
      Page Faults                       0
      Page Reclaims                     12844
      Page Swaps                        0
      Voluntary Context Switches        982
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1200

333        proc corr data=merged NOMISS plots=matrix;

334        var rate per_100000;
335        run;

NOTE: PROCEDURE CORR used (Total process time):
      real time           0.29 seconds
      user cpu time       0.07 seconds
      system cpu time     0.00 seconds
      memory              3822.81k
      OS Memory           55656.00k
      Timestamp           05/10/2019 12:20:14 AM
      Step Count                        168  Switch Count  0
      Page Faults                       0
      Page Reclaims                     420
      Page Swaps                        0
      Voluntary Context Switches        251
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           432


336
337        *The analysis shows that there is a significant relationship between the two variables;
```

```
338        *However, there appear to be (at least) two outlier variables
339        Lets remove those and see if this affects the analysis;
340        data no_outliers;
341        set merged;
342        if state = 'DC' then delete;
343        if state = 'HI' then delete;
344        *if state = 'CO' then delete;
345        *if state = 'MT' then delete;
346        *if state = 'NV' then delete;
347        run;

NOTE: There were 51 observations read from the data set WORK.MERGED.
NOTE: The data set WORK.NO_OUTLIERS has 49 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              804.59k
      OS Memory           54712.00k
      Timestamp           05/10/2019 12:20:14 AM
      Step Count                        169  Switch Count  2
      Page Faults                       0
      Page Reclaims                     121
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


348
349        *Once again we can see a significant relationship between the two
350        We can conclude that there is a significant relationship between McDonalds in a state and a state's obesity rate;
351        proc reg data=no_outliers;
352        label rate ='Obesity Rate';
353        label per_100000 ='McDonalds per 1000000';
354        model rate=per_100000;
355        run;

356

NOTE: PROCEDURE REG used (Total process time):
      real time           0.73 seconds
      user cpu time       0.23 seconds
      system cpu time     0.03 seconds
      memory              11260.50k
      OS Memory           62448.00k
      Timestamp           05/10/2019 12:20:15 AM
      Step Count                        170  Switch Count  22
      Page Faults                       0
      Page Reclaims                     12175
      Page Swaps                        0
      Voluntary Context Switches        988
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           1224

357        proc corr data=no_outliers NOMISS plots=matrix;

358        var rate per_100000;
359        run;

NOTE: PROCEDURE CORR used (Total process time):
      real time           0.24 seconds
      user cpu time       0.08 seconds
      system cpu time     0.00 seconds
      memory              3527.87k
      OS Memory           56172.00k
      Timestamp           05/10/2019 12:20:15 AM
      Step Count                        171  Switch Count  0
      Page Faults                       0
      Page Reclaims                     372
      Page Swaps                        0
      Voluntary Context Switches        251
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           448


360
361
362
363
364        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
376
```

Results: Project 3.sas

Cities with the Most McDonalds

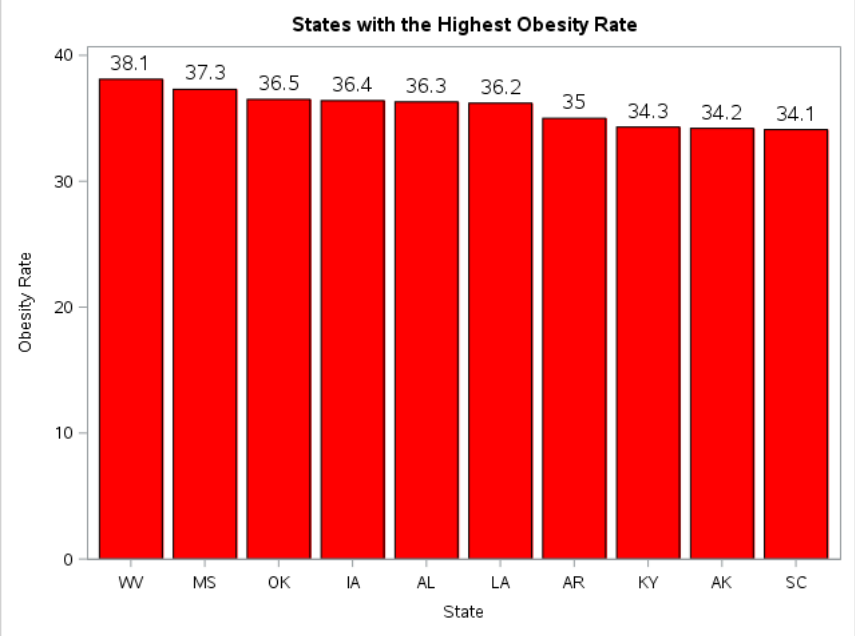
Obs	city	state	count
1	New York	NY	189
2	Houston	TX	133
3	Chicago	IL	124
4	Las Vegas	NV	83
5	Los Angeles	CA	80
6	San Antonio	TX	78
7	Dallas	TX	61
8	Phoenix	AZ	55
9	Miami	FL	55
10	Philadelphia	PA	53



States with the most McDonalds per 100,000 people

Obs	State	McDonalds/100,000
1	Kentucky	5.61722
2	West Virginia	5.48224
3	Michigan	5.46223
4	Ohio	5.44936
5	Arkansas	5.40841
6	Indiana	5.34977

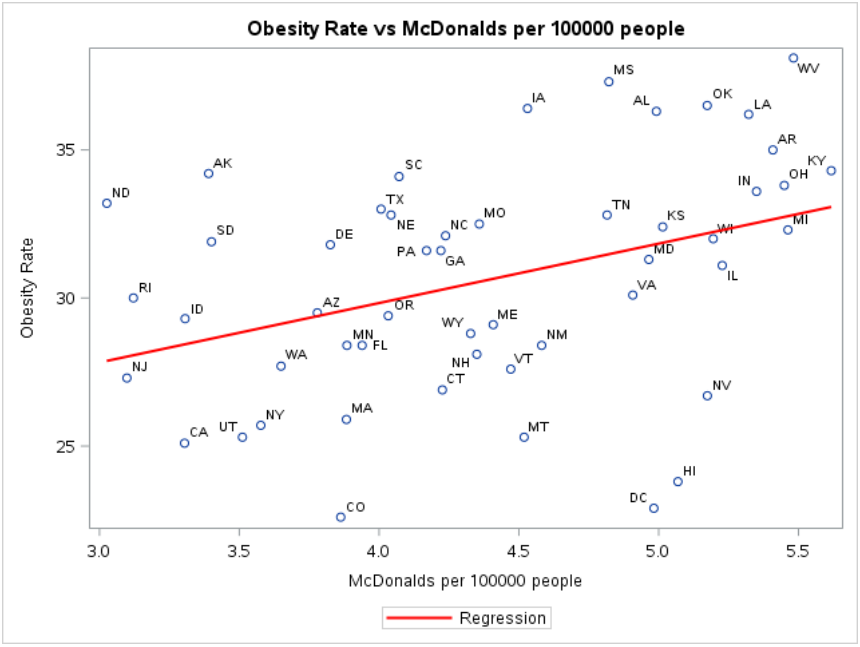
Obs	State	McDonalds/100,000
7	Louisiana	5.32191
8	Illinois	5.22719
9	Wisconsin	5.19474
10	Nevada	5.17402
11	Oklahoma	5.17362
12	Hawaii	5.06867
13	Kansas	5.01459
14	Alabama	4.99195
15	District of Columbia	4.98253
16	Maryland	4.96465
17	Virginia	4.90744
18	Mississippi	4.82165
19	Tennessee	4.81535
20	New Mexico	4.58140
21	Iowa	4.53084
22	Montana	4.51848
23	Vermont	4.47071
24	Maine	4.40824
25	Missouri	4.35815
26	New Hampshire	4.34956
27	Wyoming	4.32723
28	North Carolina	4.23744
29	Connecticut	4.22654
30	Georgia	4.22074
31	Pennsylvania	4.16958
32	South Carolina	4.07150
33	Nebraska	4.04298
34	Oregon	4.03273
35	Texas	4.00671
36	Florida	3.93909
37	Minnesota	3.88510
38	Massachusetts	3.88285
39	Colorado	3.86266
40	Delaware	3.82559
41	Arizona	3.77877
42	Washington	3.64935
43	New York	3.57687
44	Utah	3.51143
45	South Dakota	3.40045
46	Alaska	3.39012
47	Idaho	3.30634
48	California	3.30409
49	Rhode Island	3.12111
50	New Jersey	3.09816
51	North Dakota	3.02601

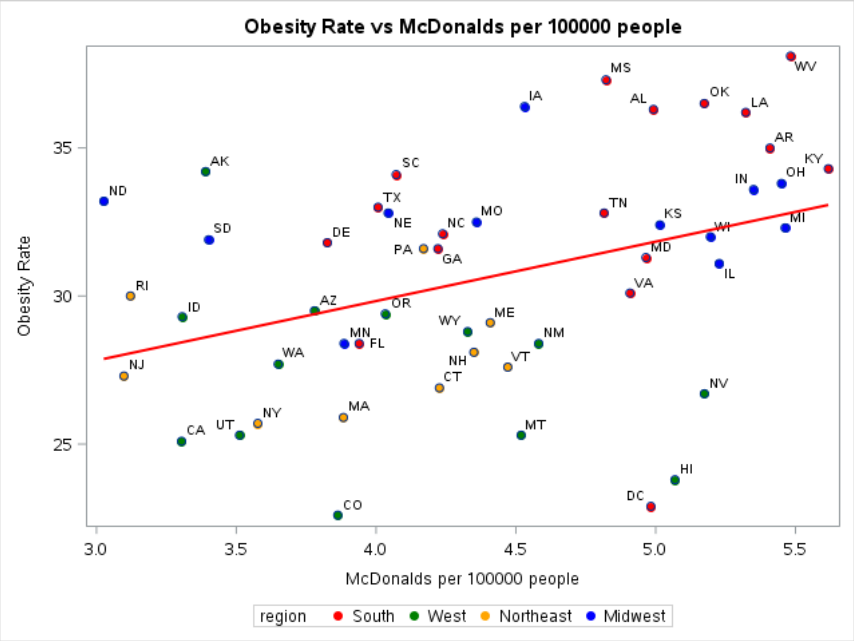


States with the Highest Obesity Rate

Obs	State	Obesity Rate (%)
1	West Virginia	38.1

Obs	State	Obesity Rate (%)
2	Mississippi	37.3
3	Oklahoma	36.5
4	Iowa	36.4
5	Alabama	36.3
6	Louisiana	36.2
7	Arkansas	35.0
8	Kentucky	34.3
9	Alaska	34.2
10	South Carolina	34.1
11	Ohio	33.8
12	Indiana	33.6
13	North Dakota	33.2
14	Texas	33.0
15	Nebraska	32.8
16	Tennessee	32.8
17	Missouri	32.5
18	Kansas	32.4
19	Michigan	32.3
20	North Carolina	32.1
21	Wisconsin	32.0
22	South Dakota	31.9
23	Delaware	31.8
24	Georgia	31.6
25	Pennsylvania	31.6
26	Maryland	31.3
27	Illinois	31.1
28	Virginia	30.1
29	Rhode Island	30.0
30	Arizona	29.5
31	Oregon	29.4
32	Idaho	29.3
33	Maine	29.1
34	Wyoming	28.8
35	Florida	28.4
36	Minnesota	28.4
37	New Mexico	28.4
38	New Hampshire	28.1
39	Washington	27.7
40	Vermont	27.6
41	New Jersey	27.3
42	Connecticut	26.9
43	Nevada	26.7
44	Massachusetts	25.9
45	New York	25.7
46	Montana	25.3
47	Utah	25.3
48	California	25.1
49	Hawaii	23.8
50	District of Columbia	22.9
51	Colorado	22.6





Obesity Rate vs McDonalds per 100000 people

The REG Procedure
Model: MODEL1
Dependent Variable: rate Obesity Rate

Number of Observations Read	51
Number of Observations Used	51

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	107.64373	107.64373	8.30	0.0059
Error	49	635.78607	12.97523		
Corrected Total	50	743.42980			

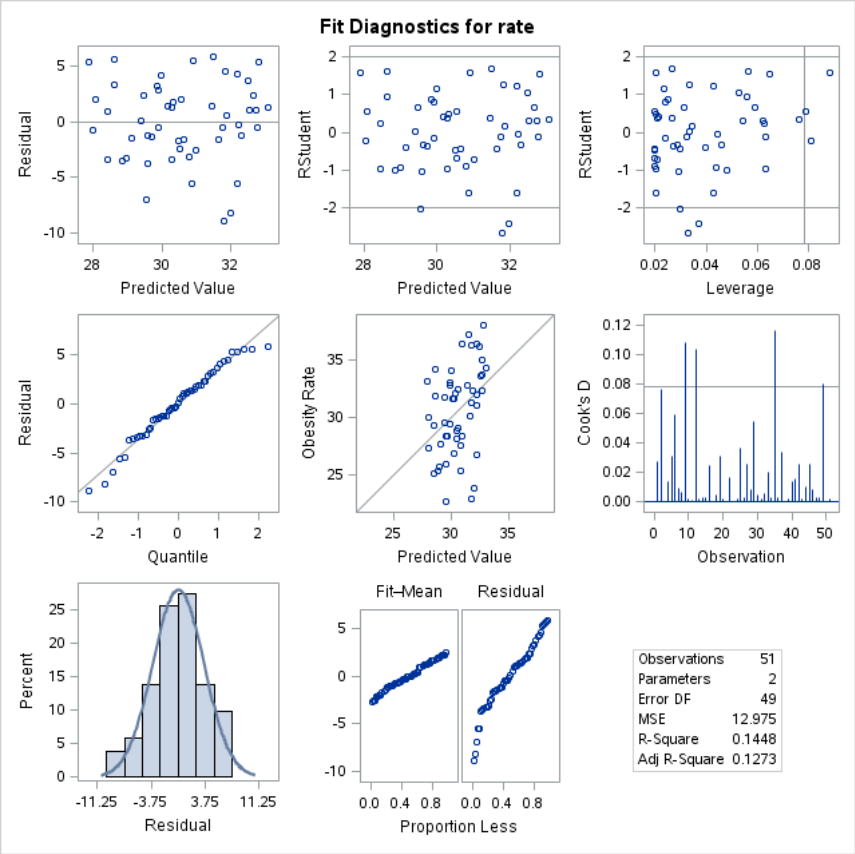
Root MSE	3.60211	R-Square	0.1448
Dependent Mean	30.59804	Adj R-Sq	0.1273
Coeff Var	11.77237		

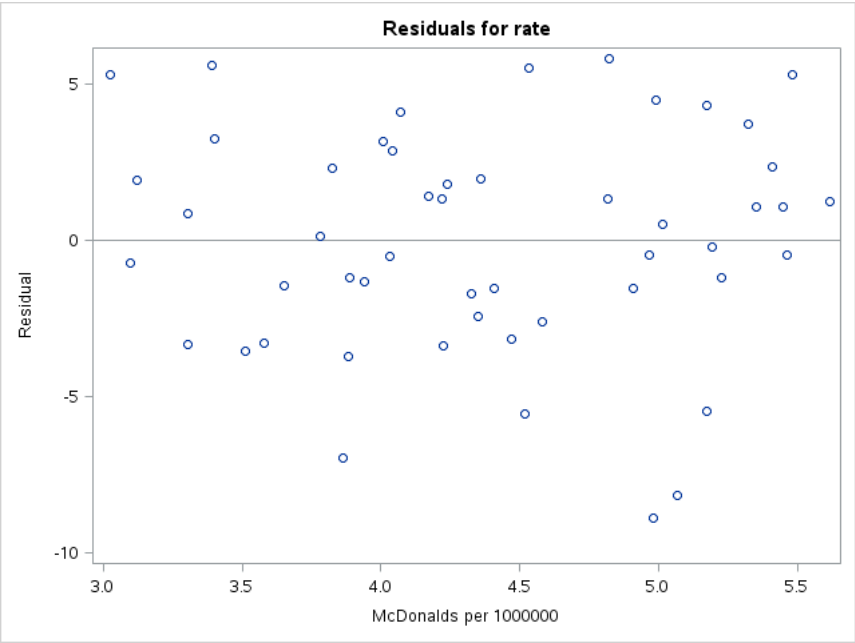
Parameter Estimates

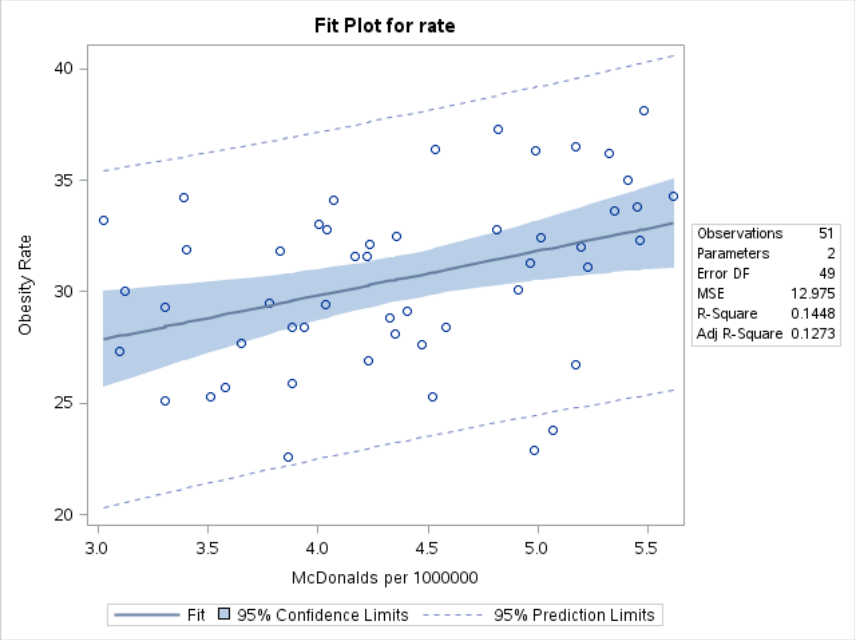
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	21.80570	3.09398	7.05	<.0001
per_100000	McDonalds per 1000000	1	2.00596	0.69644	2.88	0.0059

Obesity Rate vs McDonalds per 100000 people

The REG Procedure
Model: MODEL1
Dependent Variable: rate Obesity Rate


Fit Diagnostics for rate
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Obesity Rate
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 51
Parameters 2
Error DF 49
MSE 12.975
R-Square 0.1448
Adj R-Square 0.1273


Residuals for rate
Residual
McDonalds per 1000000



Obesity Rate vs McDonalds per 100000 people

The CORR Procedure

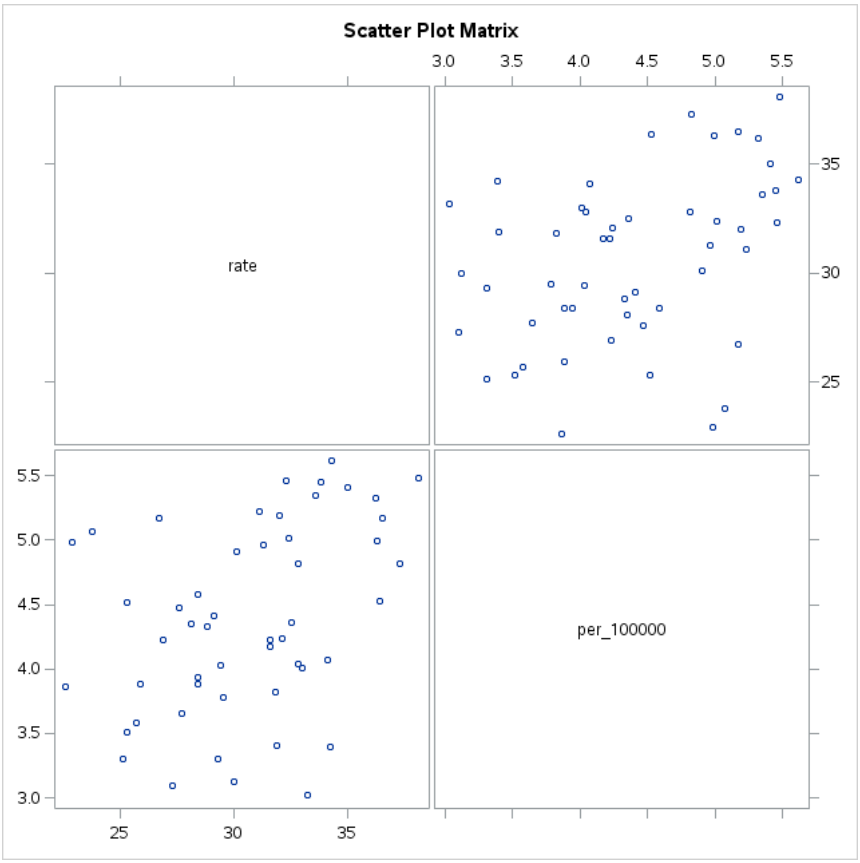
2 Variables:	rate per_100000

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
rate	51	30.59804	3.85598	1561	22.60000	38.10000
per_100000	51	4.38310	0.73145	223.53835	3.02601	5.61722

Pearson Correlation Coefficients, N = 51
Prob > |r| under H0: Rho=0

	rate	per_100000
rate	1.00000	0.38052 0.0059
per_100000	0.38052 0.0059	1.00000



Obesity Rate vs McDonalds per 100000 people

The REG Procedure
Model: MODEL1
Dependent Variable: rate Obesity Rate

Number of Observations Read	49
Number of Observations Used	49

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	154.85938	154.85938	15.20	0.0003
Error	47	478.80878	10.18742		
Corrected Total	48	633.66816			

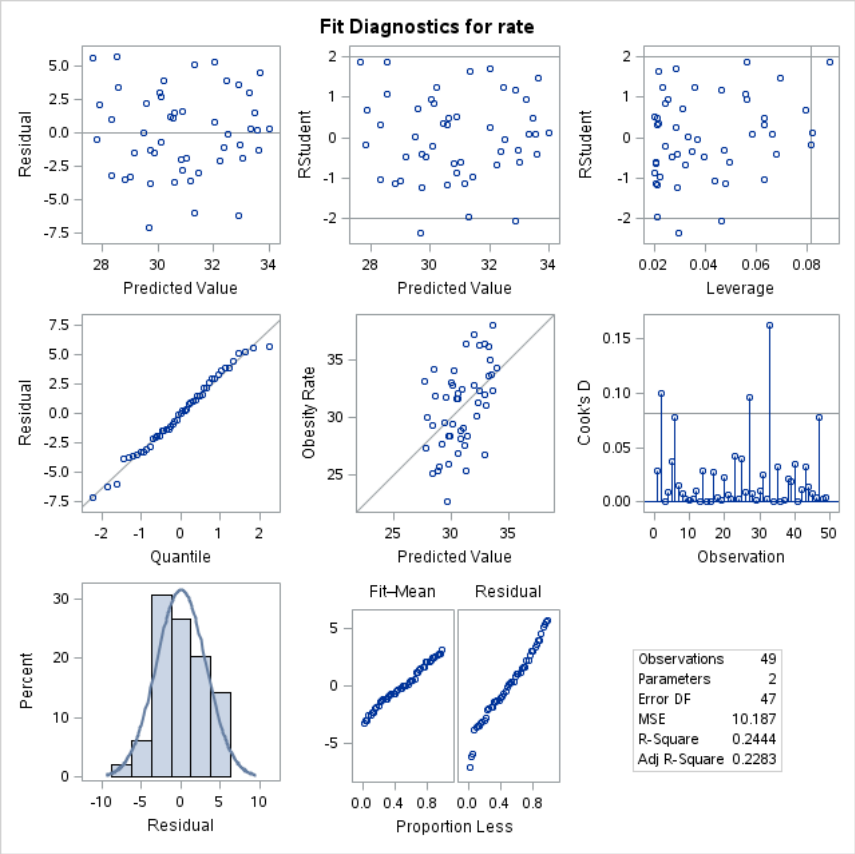
Root MSE	3.19177	R-Square	0.2444
Dependent Mean	30.89388	Adj R-Sq	0.2283
Coeff Var	10.33141		

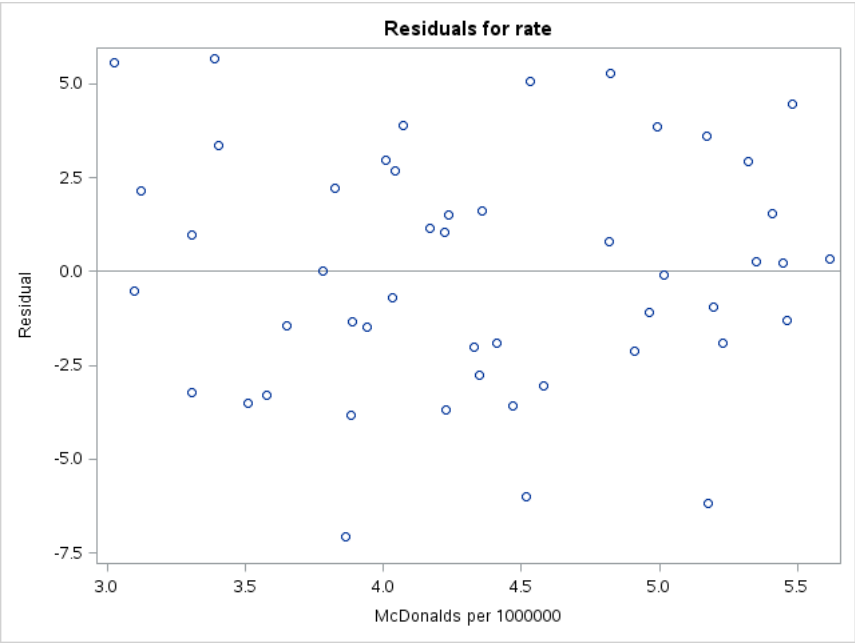
Parameter Estimates

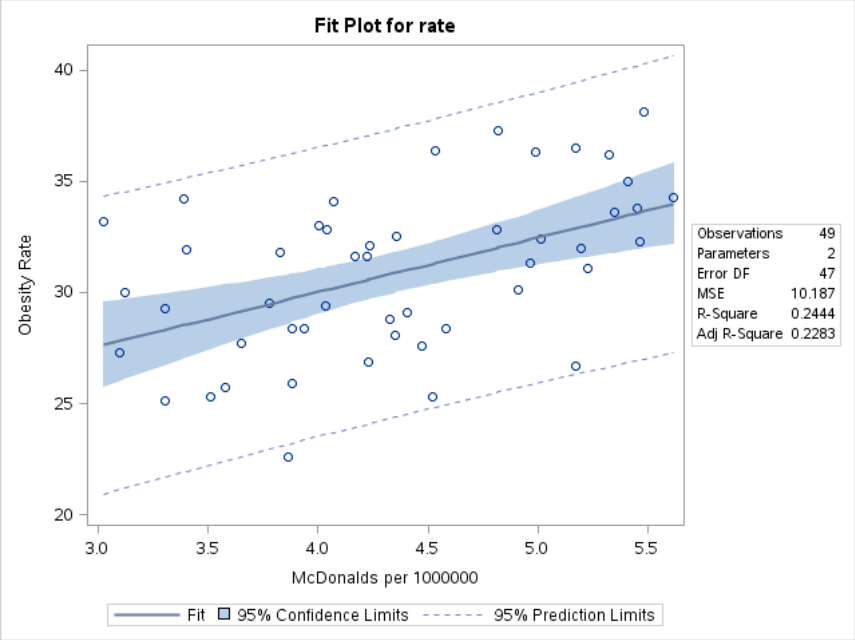
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	20.23790	2.77088	7.30	<.0001
per_100000	McDonalds per 1000000	1	2.44578	0.62731	3.90	0.0003

Obesity Rate vs McDonalds per 100000 people

The REG Procedure
Model: MODEL1
Dependent Variable: rate Obesity Rate


Fit Diagnostics for rate
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Obesity Rate
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 49
Parameters 2
Error DF 47
MSE 10.187
R-Square 0.2444
Adj R-Square 0.2283


Residuals for rate
Residual
McDonalds per 1000000



Obesity Rate vs McDonalds per 100000 people

The CORR Procedure

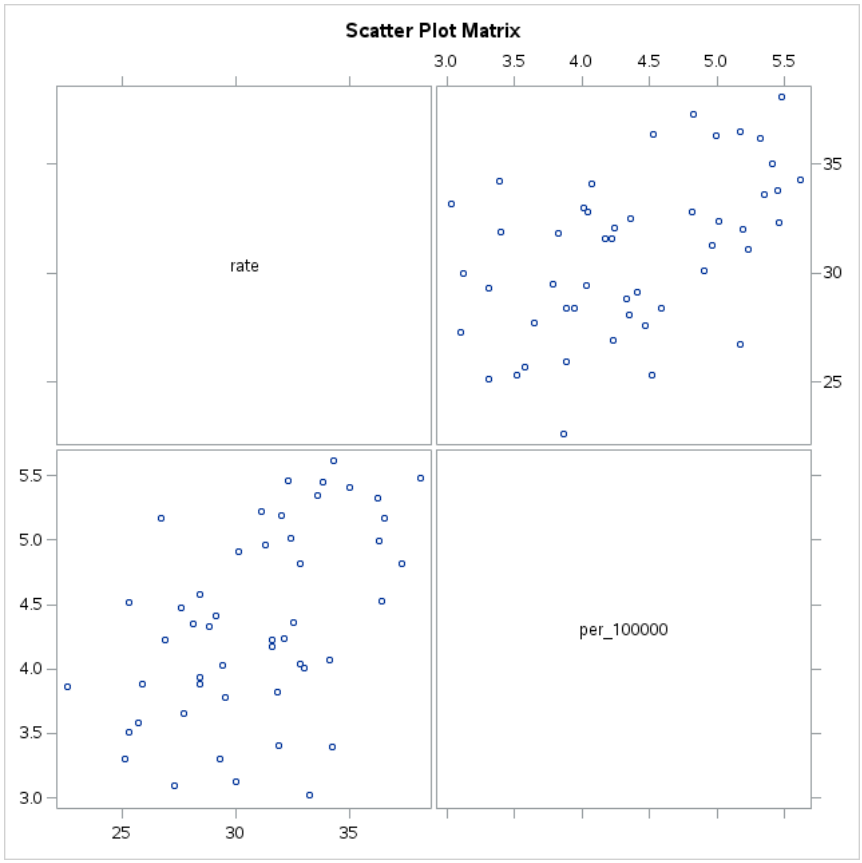
2 Variables:	rate per_100000

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
rate	49	30.89388	3.63338	1514	22.60000	38.10000
per_100000	49	4.35688	0.73440	213.48715	3.02601	5.61722

Pearson Correlation Coefficients, N = 49
Prob > |r| under H0: Rho=0

	rate	per_100000
rate	1.00000	0.49435 0.0003
per_100000	0.49435 0.0003	1.00000


Scatter Plot Matrix
3.0
3.5
4.0
4.5
5.0
5.5
rate
35
30
25
5.5
5.0
4.5
4.0
3.5
3.0
per_100000
25
30
35