BDA_Lab02_SparkSQL

Saewon Jun(saeju204), Anubhav Dikshit(anudi287)

There are two ways to write queries in Spark SQL : **1)Using built-in API functions** or **2)running SQL-like queries**. In this lab, we will use built-in API functinos for all the questions and write a regular SQL qeury only for the second question.

1.What are the lowest and highest temperatures measured each year for the period 1950-2014. Provide the lists sorted in the descending order with respect to the maximum temperature.

year, station with the min, maxValue ORDER BY minValue DESC(value=Temp) year, station with the max, maxValue ORDER BY maxValue DESC(value=Temp)

```python
from pyspark import SparkContext
from pyspark.sql import functions as F
from pyspark.sql import SQLContext, Row

#SparkContext is required when we want to execute operations in a cluster.
#SparkContext tells Spark how and where to access a cluster. And the first step is to connect with Apac

##create the SparkContext, as well as sqlContext
sc = SparkContext(appName="BDA2_Q1")
sqlContext = SQLContext(sc)

##read the data and preprocess the data
data = sc.textFile("/user/x_saeju/data/temperature-readings.csv").cache()
data = data.map(lambda line: line.split(";"))
data = data.filter(lambda x:(int(x[1][0:4]) >= 1950 and int(x[1][0:4]) <=2014))
data = data.map(lambda x: Row(station=x[0], date=x[1], year=x[1].split("-")[0], time=x[2],
                temp=float(x[3]), quality=x[4]))
                #Row() can be used to create a row object by using named arguments, the fields will be
                #you can access the data by i.e)data.station, data.time

q1Data = sqlContext.createDataFrame(data)
q1Data.registerTempTable("Q1_dataframe")
 # registerTempTable(): does not create a "cached" in-memory table, but rather an alias or a reference
 #Every time you query the temp table you are actually re-executing the DAG unless you explicitly cache

 #cf) saveAsTable(): creates a permanent, physical table stored in S3 using the Parquet format.
 #This table is accessible to all clusters including the dashboard cluster.


##For the lowest temperatures measured each year for the period 1950-2014
minTemp = q1Data.groupBy("year").agg(F.min("temp").alias("temp")).orderBy(["temp"], ascending=[0])
        #agg():aggregate on the entire DataFrame without groups
        #alias():returns a new DataFrame with an alias set.
minTemp = minTemp.join(q1Data.select("station","year","temp"),["year","temp"], 'left_outer')
minTemp = minTemp.orderBy(["temp"], ascending=[0])
minTemp = minTemp.rdd.repartition(1).sortBy(ascending=False, keyfunc=lambda x: x[1])
minTemp.repartition(1).saveAsTextFile("/user/x_saeju/results/Q1_lowestTemp")
```

```
##For the highest temperatures measured each year for the period 1950-2014
maxTemp = q1Data.groupBy("year").agg(F.max("temp").alias("temp")).orderBy(["temp"], ascending=[0])
maxTemp = maxTemp.join(q1Data.select("station","year","temp"),["year","temp"], 'left_outer')
maxTemp = maxTemp.orderBy(["temp"], ascending=[0])
maxTemp = maxTemp.rdd.repartition(1).sortBy(ascending=False, keyfunc=lambda x: x[1])
maxTemp.repartition(1).saveAsTextFile("/user/x_saeju/results/Q1_highestTemp")
```

Q1_highestTemp

```
Row(year=u'1975', temp=36.1, station=u'86200')
Row(year=u'1992', temp=35.4, station=u'63600')
Row(year=u'1994', temp=34.7, station=u'117160')
Row(year=u'2014', temp=34.4, station=u'96560')
Row(year=u'2010', temp=34.4, station=u'75250')
Row(year=u'1989', temp=33.9, station=u'63050')
Row(year=u'1982', temp=33.8, station=u'94050')
Row(year=u'1968', temp=33.7, station=u'137100')
Row(year=u'1966', temp=33.5, station=u'151640')
Row(year=u'1983', temp=33.3, station=u'98210')
Row(year=u'2002', temp=33.3, station=u'78290')
Row(year=u'2002', temp=33.3, station=u'78290')
Row(year=u'1986', temp=33.2, station=u'76470')
Row(year=u'1970', temp=33.2, station=u'103080')
Row(year=u'1956', temp=33.0, station=u'145340')
Row(year=u'2000', temp=33.0, station=u'62400')
Row(year=u'1959', temp=32.8, station=u'65160')
Row(year=u'2006', temp=32.7, station=u'75240')
Row(year=u'1991', temp=32.7, station=u'137040')
Row(year=u'1988', temp=32.6, station=u'102540')
Row(year=u'2011', temp=32.5, station=u'172770')
Row(year=u'1999', temp=32.4, station=u'98210')
Row(year=u'2008', temp=32.2, station=u'102390')
Row(year=u'2008', temp=32.2, station=u'82090')
Row(year=u'2008', temp=32.2, station=u'82090')
Row(year=u'2008', temp=32.2, station=u'95130')
Row(year=u'2003', temp=32.2, station=u'136420')
Row(year=u'1953', temp=32.2, station=u'65160')
Row(year=u'2007', temp=32.2, station=u'86420')
Row(year=u'1955', temp=32.2, station=u'97260')
Row(year=u'1973', temp=32.2, station=u'71470')
Row(year=u'2005', temp=32.1, station=u'107140')
Row(year=u'1979', temp=32.0, station=u'63600')
Row(year=u'1969', temp=32.0, station=u'97260')
Row(year=u'1969', temp=32.0, station=u'71470')
Row(year=u'2001', temp=31.9, station=u'62400')
Row(year=u'1977', temp=31.8, station=u'94180')
Row(year=u'1997', temp=31.8, station=u'76420')
Row(year=u'1997', temp=31.8, station=u'74180')
Row(year=u'2013', temp=31.6, station=u'98210')
Row(year=u'2009', temp=31.5, station=u'81370')
Row(year=u'2012', temp=31.3, station=u'54990')
Row(year=u'1964', temp=31.2, station=u'76430')
Row(year=u'1971', temp=31.2, station=u'65130')
```

```
Row(year=u'1972', temp=31.2, station=u'137100')
Row(year=u'1972', temp=31.2, station=u'97260')
Row(year=u'1972', temp=31.2, station=u'98200')
Row(year=u'1976', temp=31.1, station=u'75040')
Row(year=u'1961', temp=31.0, station=u'76000')
Row(year=u'1963', temp=31.0, station=u'62410')
Row(year=u'1958', temp=30.8, station=u'117160')
Row(year=u'1995', temp=30.8, station=u'76420')
Row(year=u'1978', temp=30.8, station=u'71470')
Row(year=u'1996', temp=30.8, station=u'96140')
Row(year=u'1996', temp=30.8, station=u'98210')
Row(year=u'1974', temp=30.6, station=u'167710')
Row(year=u'1974', temp=30.6, station=u'183980')
Row(year=u'1954', temp=30.5, station=u'53650')
Row(year=u'1980', temp=30.4, station=u'161790')
Row(year=u'1980', temp=30.4, station=u'92400')
Row(year=u'1952', temp=30.4, station=u'106100')
Row(year=u'2004', temp=30.2, station=u'117160')
Row(year=u'2004', temp=30.2, station=u'117160')
Row(year=u'1990', temp=30.2, station=u'54330')
Row(year=u'1990', temp=30.2, station=u'76000')
Row(year=u'1990', temp=30.2, station=u'63340')
Row(year=u'1985', temp=29.8, station=u'85450')
Row(year=u'1957', temp=29.8, station=u'75040')
Row(year=u'1981', temp=29.7, station=u'65160')
Row(year=u'1993', temp=29.7, station=u'132030')
Row(year=u'1987', temp=29.6, station=u'105410')
Row(year=u'1987', temp=29.6, station=u'98210')
Row(year=u'1987', temp=29.6, station=u'96310')
Row(year=u'1984', temp=29.5, station=u'105370')
Row(year=u'1967', temp=29.5, station=u'75040')
Row(year=u'1960', temp=29.4, station=u'173810')
Row(year=u'1950', temp=29.4, station=u'75040')
Row(year=u'1998', temp=29.2, station=u'63600')
Row(year=u'1965', temp=28.5, station=u'116500')
Row(year=u'1951', temp=28.5, station=u'75040')
Row(year=u'1962', temp=27.4, station=u'76380')
Row(year=u'1962', temp=27.4, station=u'86200')
```

Q1_lowestTemp

```
Row(year=u'1990', temp=-35.0, station=u'147270')
Row(year=u'1990', temp=-35.0, station=u'166870')
Row(year=u'1952', temp=-35.5, station=u'192830')
Row(year=u'1974', temp=-35.6, station=u'166870')
Row(year=u'1974', temp=-35.6, station=u'179950')
Row(year=u'1954', temp=-36.0, station=u'113410')
Row(year=u'1992', temp=-36.1, station=u'179960')
Row(year=u'1975', temp=-37.0, station=u'157860')
Row(year=u'1972', temp=-37.5, station=u'167860')
Row(year=u'1995', temp=-37.6, station=u'182910')
Row(year=u'2000', temp=-37.6, station=u'169860')
Row(year=u'1957', temp=-37.8, station=u'159970')
```

```
Row(year=u'1983', temp=-38.2, station=u'191900')
Row(year=u'1989', temp=-38.2, station=u'166870')
Row(year=u'1953', temp=-38.4, station=u'183760')
Row(year=u'2009', temp=-38.5, station=u'179960')
Row(year=u'1993', temp=-39.0, station=u'191900')
Row(year=u'1984', temp=-39.2, station=u'191900')
Row(year=u'1984', temp=-39.2, station=u'123480')
Row(year=u'1973', temp=-39.3, station=u'166870')
Row(year=u'2008', temp=-39.3, station=u'179960')
Row(year=u'2008', temp=-39.3, station=u'179960')
Row(year=u'1991', temp=-39.3, station=u'179960')
Row(year=u'2005', temp=-39.4, station=u'155790')
Row(year=u'1964', temp=-39.5, station=u'166810')
Row(year=u'1961', temp=-39.5, station=u'181900')
Row(year=u'1970', temp=-39.6, station=u'179950')
Row(year=u'2004', temp=-39.7, station=u'166940')
Row(year=u'1988', temp=-39.9, station=u'170790')
Row(year=u'1960', temp=-40.0, station=u'155910')
Row(year=u'1960', temp=-40.0, station=u'160790')
Row(year=u'1960', temp=-40.0, station=u'167710')
Row(year=u'1997', temp=-40.2, station=u'179960')
Row(year=u'1994', temp=-40.5, station=u'179960')
Row(year=u'2006', temp=-40.6, station=u'169860')
Row(year=u'2013', temp=-40.7, station=u'179960')
Row(year=u'2007', temp=-40.7, station=u'169860')
Row(year=u'2007', temp=-40.7, station=u'169860')
Row(year=u'1963', temp=-41.0, station=u'181900')
Row(year=u'1955', temp=-41.2, station=u'160790')
Row(year=u'2003', temp=-41.5, station=u'179960')
Row(year=u'1969', temp=-41.5, station=u'181900')
Row(year=u'1996', temp=-41.7, station=u'155790')
Row(year=u'2010', temp=-41.7, station=u'191910')
Row(year=u'1950', temp=-42.0, station=u'155910')
Row(year=u'1962', temp=-42.0, station=u'181900')
Row(year=u'1962', temp=-42.0, station=u'181900')
Row(year=u'1951', temp=-42.0, station=u'155910')
Row(year=u'1968', temp=-42.0, station=u'179950')
Row(year=u'2011', temp=-42.0, station=u'179960')
Row(year=u'1982', temp=-42.2, station=u'113410')
Row(year=u'1976', temp=-42.2, station=u'192830')
Row(year=u'2002', temp=-42.2, station=u'169860')
Row(year=u'1977', temp=-42.5, station=u'179950')
Row(year=u'2014', temp=-42.5, station=u'192840')
Row(year=u'1998', temp=-42.7, station=u'169860')
Row(year=u'1998', temp=-42.7, station=u'169860')
Row(year=u'2012', temp=-42.7, station=u'191910')
Row(year=u'1958', temp=-43.0, station=u'159970')
Row(year=u'1985', temp=-43.4, station=u'159970')
Row(year=u'1985', temp=-43.4, station=u'166870')
Row(year=u'1959', temp=-43.6, station=u'159970')
Row(year=u'1979', temp=-44.0, station=u'112170')
Row(year=u'1981', temp=-44.0, station=u'166870')
Row(year=u'1981', temp=-44.0, station=u'166870')
```

```
Row(year=u'1965', temp=-44.0, station=u'189780')
Row(year=u'2001', temp=-44.0, station=u'112530')
Row(year=u'1986', temp=-44.2, station=u'167860')
Row(year=u'1971', temp=-44.3, station=u'166870')
Row(year=u'1956', temp=-45.0, station=u'160790')
Row(year=u'1980', temp=-45.0, station=u'191900')
Row(year=u'1980', temp=-45.0, station=u'191900')
Row(year=u'1967', temp=-45.4, station=u'166870')
Row(year=u'1987', temp=-47.3, station=u'123480')
Row(year=u'1978', temp=-47.7, station=u'155940')
Row(year=u'1999', temp=-49.0, station=u'192830')
Row(year=u'1999', temp=-49.0, station=u'192830')
Row(year=u'1966', temp=-49.4, station=u'179950')
```

2.Count the number of readings for each month in the period of 1950-2014 which are higer than 10 degrees. Then repeat this exercise taking only distinct readings from each station.

year, month, value ORDER BY value DESC(value=count)

```
from pyspark import SparkContext
from pyspark.sql import functions as F
from pyspark.sql import SQLContext, Row

#sc.stop()
sc = SparkContext(appName="BDA2_Q2")
sqlContext = SQLContext(sc)

#Read the data
data = sc.textFile("/user/x_saeju/data/temperature-readings.csv").cache()
data = data.map(lambda line: line.split(";"))
data = data.filter(lambda x:(int(x[1][0:4])>=1950 and int(x[1][0:4])<=2014))
data = data.map(lambda x: Row(station=x[0], date=x[1], year=x[1].split("-")[0],
                              month=x[1].split("-")[1], time=x[2], temp=float(x[3]), quality=x[4]))

q2data = sqlContext.createDataFrame(data)
q2data.registerTempTable("Q2_dataframe")

###API method
Q2_API = q2data.filter(q2data["temp"]>10).groupby(["year","month"]).agg(F.countDistinct("station"))
Q2_API = Q2_API.rdd.repartition(1).sortBy(ascending=False, keyfunc=lambda x:x[2])
Q2_API.repartition(1).saveAsTextFile("/user/x_saeju/results/Q2_API")

###SQL method
Q2_SQL = sqlContext.sql("SELECT year, month, count(DISTINCT(station)) as count
                         FROM Q2_dataframe WHERE temp>10
                         GROUP BY year, month ORDER BY count DESC")
Q2_SQL = Q2_SQL.rdd.repartition(1).sortyBy(ascending=False, keyfunc=lambda x:x[2])
Q2_SQL.repartition(1).saveAsTextFile("/user/x_saeju/results/Q2_SQL")
```

API methods

```
Row(year=u'1972', month=u'10', count(station)=378)
Row(year=u'1973', month=u'05', count(station)=377)
Row(year=u'1973', month=u'06', count(station)=377)
Row(year=u'1972', month=u'08', count(station)=376)
Row(year=u'1973', month=u'09', count(station)=376)
Row(year=u'1972', month=u'06', count(station)=375)
Row(year=u'1972', month=u'09', count(station)=375)
Row(year=u'1971', month=u'08', count(station)=375)
Row(year=u'1972', month=u'05', count(station)=375)
Row(year=u'1972', month=u'07', count(station)=374)
Row(year=u'1971', month=u'06', count(station)=374)
Row(year=u'1971', month=u'09', count(station)=374)
Row(year=u'1973', month=u'08', count(station)=373)
Row(year=u'1971', month=u'05', count(station)=373)
Row(year=u'1974', month=u'06', count(station)=372)
Row(year=u'1974', month=u'08', count(station)=372)
Row(year=u'1973', month=u'07', count(station)=370)
Row(year=u'1974', month=u'05', count(station)=370)
Row(year=u'1974', month=u'09', count(station)=370)
Row(year=u'1970', month=u'08', count(station)=370)
Row(year=u'1971', month=u'07', count(station)=370)
Row(year=u'1975', month=u'09', count(station)=369)
Row(year=u'1970', month=u'06', count(station)=369)
Row(year=u'1970', month=u'09', count(station)=369)
Row(year=u'1976', month=u'05', count(station)=369)
Row(year=u'1975', month=u'06', count(station)=368)
Row(year=u'1976', month=u'06', count(station)=368)
Row(year=u'1975', month=u'05', count(station)=367)
Row(year=u'1975', month=u'08', count(station)=367)
Row(year=u'1970', month=u'05', count(station)=366)
Row(year=u'1976', month=u'09', count(station)=365)
Row(year=u'1977', month=u'06', count(station)=364)
Row(year=u'1967', month=u'05', count(station)=363)
Row(year=u'1976', month=u'08', count(station)=363)
Row(year=u'1974', month=u'07', count(station)=362)
Row(year=u'1970', month=u'07', count(station)=362)
Row(year=u'1967', month=u'09', count(station)=361)
Row(year=u'1966', month=u'06', count(station)=360)
Row(year=u'1966', month=u'09', count(station)=360)
Row(year=u'1966', month=u'08', count(station)=359)
Row(year=u'1967', month=u'06', count(station)=359)
Row(year=u'1969', month=u'09', count(station)=359)
Row(year=u'1978', month=u'09', count(station)=358)
Row(year=u'1967', month=u'08', count(station)=358)
Row(year=u'1975', month=u'07', count(station)=358)
Row(year=u'1965', month=u'09', count(station)=358)
Row(year=u'1968', month=u'06', count(station)=357)
Row(year=u'1968', month=u'08', count(station)=357)
Row(year=u'1969', month=u'08', count(station)=357)
Row(year=u'1968', month=u'09', count(station)=356)
Row(year=u'1976', month=u'07', count(station)=356)
Row(year=u'1968', month=u'05', count(station)=355)
Row(year=u'1965', month=u'06', count(station)=355)
```

```
Row(year=u'1966', month=u'05', count(station)=354)
Row(year=u'1978', month=u'06', count(station)=354)
Row(year=u'1979', month=u'05', count(station)=354)
Row(year=u'1965', month=u'08', count(station)=354)
Row(year=u'1977', month=u'08', count(station)=354)
Row(year=u'1968', month=u'07', count(station)=353)
Row(year=u'1977', month=u'09', count(station)=353)
Row(year=u'1966', month=u'07', count(station)=352)
Row(year=u'1978', month=u'05', count(station)=352)
Row(year=u'1969', month=u'06', count(station)=352)
Row(year=u'1967', month=u'07', count(station)=351)
Row(year=u'1979', month=u'06', count(station)=351)
Row(year=u'1979', month=u'09', count(station)=351)
Row(year=u'1977', month=u'05', count(station)=351)
Row(year=u'1978', month=u'08', count(station)=350)
Row(year=u'1977', month=u'07', count(station)=350)
Row(year=u'1973', month=u'10', count(station)=349)
Row(year=u'1969', month=u'07', count(station)=349)
Row(year=u'1965', month=u'07', count(station)=349)
Row(year=u'1971', month=u'10', count(station)=347)
Row(year=u'1969', month=u'10', count(station)=346)
Row(year=u'1996', month=u'06', count(station)=345)
Row(year=u'1979', month=u'07', count(station)=345)
Row(year=u'1970', month=u'10', count(station)=345)
Row(year=u'1974', month=u'04', count(station)=344)
Row(year=u'1965', month=u'05', count(station)=344)
Row(year=u'1978', month=u'07', count(station)=343)
Row(year=u'1996', month=u'05', count(station)=342)
Row(year=u'1996', month=u'07', count(station)=342)
Row(year=u'1996', month=u'08', count(station)=341)
Row(year=u'1996', month=u'09', count(station)=340)
Row(year=u'1978', month=u'10', count(station)=340)
Row(year=u'1979', month=u'08', count(station)=340)
Row(year=u'1997', month=u'09', count(station)=340)
Row(year=u'1975', month=u'10', count(station)=340)
Row(year=u'1982', month=u'06', count(station)=339)
Row(year=u'1980', month=u'09', count(station)=338)
Row(year=u'1997', month=u'06', count(station)=338)
Row(year=u'1980', month=u'05', count(station)=337)
Row(year=u'1997', month=u'08', count(station)=337)
Row(year=u'1981', month=u'05', count(station)=337)
Row(year=u'1983', month=u'06', count(station)=337)
Row(year=u'1983', month=u'05', count(station)=336)
Row(year=u'1969', month=u'05', count(station)=335)
Row(year=u'1981', month=u'09', count(station)=335)
Row(year=u'1965', month=u'10', count(station)=335)
Row(year=u'1981', month=u'08', count(station)=334)
Row(year=u'1982', month=u'09', count(station)=334)
Row(year=u'1984', month=u'05', count(station)=333)
Row(year=u'1997', month=u'07', count(station)=333)
Row(year=u'1980', month=u'06', count(station)=332)
Row(year=u'1983', month=u'09', count(station)=332)
Row(year=u'1981', month=u'06', count(station)=331)
```

```
Row(year=u'1980', month=u'08', count(station)=330)
Row(year=u'1982', month=u'05', count(station)=330)
Row(year=u'1999', month=u'06', count(station)=330)
Row(year=u'1983', month=u'08', count(station)=330)
Row(year=u'1981', month=u'07', count(station)=329)
Row(year=u'1999', month=u'07', count(station)=329)
Row(year=u'1999', month=u'09', count(station)=328)
Row(year=u'1984', month=u'09', count(station)=327)
Row(year=u'1985', month=u'09', count(station)=327)
Row(year=u'1999', month=u'08', count(station)=327)
Row(year=u'2002', month=u'06', count(station)=326)
Row(year=u'1998', month=u'06', count(station)=326)
Row(year=u'1998', month=u'07', count(station)=326)
Row(year=u'1998', month=u'08', count(station)=326)
Row(year=u'1998', month=u'09', count(station)=326)
Row(year=u'1982', month=u'08', count(station)=326)
Row(year=u'1985', month=u'05', count(station)=325)
Row(year=u'1981', month=u'10', count(station)=325)
Row(year=u'1999', month=u'05', count(station)=325)
Row(year=u'2000', month=u'08', count(station)=325)
Row(year=u'1984', month=u'06', count(station)=324)
Row(year=u'2002', month=u'05', count(station)=324)
Row(year=u'2002', month=u'07', count(station)=324)
Row(year=u'1985', month=u'06', count(station)=324)
Row(year=u'1967', month=u'10', count(station)=324)
Row(year=u'1980', month=u'07', count(station)=324)
Row(year=u'2001', month=u'06', count(station)=324)
Row(year=u'2001', month=u'07', count(station)=324)
Row(year=u'2002', month=u'09', count(station)=323)
Row(year=u'2003', month=u'06', count(station)=323)
Row(year=u'1986', month=u'09', count(station)=323)
Row(year=u'1987', month=u'06', count(station)=323)
Row(year=u'1987', month=u'09', count(station)=323)
Row(year=u'2000', month=u'05', count(station)=323)
Row(year=u'2001', month=u'08', count(station)=323)
Row(year=u'2002', month=u'08', count(station)=322)
Row(year=u'1968', month=u'04', count(station)=322)
Row(year=u'1998', month=u'05', count(station)=322)
Row(year=u'2000', month=u'09', count(station)=322)
Row(year=u'1988', month=u'06', count(station)=322)
Row(year=u'2001', month=u'09', count(station)=322)
Row(year=u'1984', month=u'10', count(station)=321)
Row(year=u'1991', month=u'06', count(station)=321)
Row(year=u'2003', month=u'05', count(station)=321)
Row(year=u'2003', month=u'07', count(station)=321)
Row(year=u'2004', month=u'05', count(station)=321)
Row(year=u'2004', month=u'09', count(station)=321)
Row(year=u'1982', month=u'07', count(station)=321)
Row(year=u'2000', month=u'06', count(station)=321)
Row(year=u'2003', month=u'08', count(station)=320)
Row(year=u'2003', month=u'09', count(station)=320)
Row(year=u'2010', month=u'06', count(station)=320)
Row(year=u'2004', month=u'08', count(station)=320)
```

```
Row(year=u'1987', month=u'05', count(station)=320)
Row(year=u'1987', month=u'08', count(station)=320)
Row(year=u'2000', month=u'07', count(station)=320)
Row(year=u'1988', month=u'05', count(station)=320)
Row(year=u'1997', month=u'05', count(station)=319)
Row(year=u'2010', month=u'05', count(station)=319)
Row(year=u'2004', month=u'06', count(station)=319)
Row(year=u'2004', month=u'07', count(station)=319)
Row(year=u'1987', month=u'07', count(station)=319)
Row(year=u'2011', month=u'07', count(station)=319)
Row(year=u'1983', month=u'07', count(station)=319)
Row(year=u'1989', month=u'05', count(station)=318)
Row(year=u'1995', month=u'08', count(station)=318)
Row(year=u'1984', month=u'08', count(station)=318)
Row(year=u'2010', month=u'07', count(station)=318)
Row(year=u'2010', month=u'08', count(station)=318)
Row(year=u'2011', month=u'06', count(station)=318)
Row(year=u'1985', month=u'10', count(station)=317)
Row(year=u'1986', month=u'05', count(station)=317)
Row(year=u'1987', month=u'10', count(station)=317)
Row(year=u'1989', month=u'08', count(station)=316)
Row(year=u'1989', month=u'09', count(station)=316)
Row(year=u'1990', month=u'08', count(station)=316)
Row(year=u'2008', month=u'05', count(station)=316)
Row(year=u'2008', month=u'06', count(station)=316)
Row(year=u'2010', month=u'09', count(station)=316)
Row(year=u'2011', month=u'08', count(station)=316)
Row(year=u'1988', month=u'08', count(station)=316)
Row(year=u'1989', month=u'06', count(station)=315)
Row(year=u'1995', month=u'09', count(station)=315)
Row(year=u'1991', month=u'08', count(station)=315)
Row(year=u'2011', month=u'05', count(station)=315)
Row(year=u'2011', month=u'09', count(station)=315)
Row(year=u'1988', month=u'09', count(station)=315)
Row(year=u'1982', month=u'10', count(station)=315)
Row(year=u'2000', month=u'10', count(station)=315)
Row(year=u'2001', month=u'05', count(station)=315)
Row(year=u'2008', month=u'07', count(station)=314)
Row(year=u'1991', month=u'07', count(station)=314)
Row(year=u'1985', month=u'08', count(station)=314)
Row(year=u'1986', month=u'06', count(station)=314)
Row(year=u'1984', month=u'04', count(station)=313)
Row(year=u'1990', month=u'06', count(station)=313)
Row(year=u'2008', month=u'08', count(station)=313)
Row(year=u'1989', month=u'07', count(station)=312)
Row(year=u'1984', month=u'07', count(station)=312)
Row(year=u'1990', month=u'09', count(station)=312)
Row(year=u'2009', month=u'07', count(station)=312)
Row(year=u'2009', month=u'09', count(station)=312)
Row(year=u'1991', month=u'09', count(station)=311)
Row(year=u'2009', month=u'08', count(station)=311)
Row(year=u'1963', month=u'08', count(station)=311)
Row(year=u'1992', month=u'05', count(station)=311)
```

```
Row(year=u'1992', month=u'07', count(station)=311)
Row(year=u'2005', month=u'06', count(station)=311)
Row(year=u'2009', month=u'06', count(station)=310)
Row(year=u'1963', month=u'05', count(station)=310)
Row(year=u'1963', month=u'06', count(station)=310)
Row(year=u'1963', month=u'09', count(station)=310)
Row(year=u'1992', month=u'06', count(station)=310)
Row(year=u'1992', month=u'08', count(station)=310)
Row(year=u'2012', month=u'06', count(station)=310)
Row(year=u'2006', month=u'06', count(station)=310)
Row(year=u'2012', month=u'07', count(station)=310)
Row(year=u'2006', month=u'07', count(station)=310)
Row(year=u'2012', month=u'08', count(station)=310)
Row(year=u'2006', month=u'08', count(station)=309)
Row(year=u'2006', month=u'09', count(station)=309)
Row(year=u'1990', month=u'07', count(station)=308)
Row(year=u'2009', month=u'05', count(station)=308)
Row(year=u'1963', month=u'07', count(station)=308)
Row(year=u'2012', month=u'05', count(station)=308)
Row(year=u'1990', month=u'05', count(station)=307)
Row(year=u'2005', month=u'07', count(station)=307)
Row(year=u'2005', month=u'09', count(station)=307)
Row(year=u'2006', month=u'05', count(station)=307)
Row(year=u'2005', month=u'08', count(station)=306)
Row(year=u'1988', month=u'07', count(station)=306)
Row(year=u'2012', month=u'09', count(station)=306)
Row(year=u'1962', month=u'10', count(station)=305)
Row(year=u'2007', month=u'07', count(station)=304)
Row(year=u'1985', month=u'07', count(station)=304)
Row(year=u'1992', month=u'09', count(station)=304)
Row(year=u'1986', month=u'08', count(station)=304)
Row(year=u'1964', month=u'05', count(station)=304)
Row(year=u'1965', month=u'04', count(station)=304)
Row(year=u'1977', month=u'10', count(station)=303)
Row(year=u'2007', month=u'06', count(station)=303)
Row(year=u'1964', month=u'06', count(station)=303)
Row(year=u'1994', month=u'06', count(station)=303)
Row(year=u'2013', month=u'06', count(station)=302)
Row(year=u'2007', month=u'08', count(station)=302)
Row(year=u'1995', month=u'10', count(station)=302)
Row(year=u'2008', month=u'09', count(station)=302)
Row(year=u'2005', month=u'05', count(station)=302)
Row(year=u'1994', month=u'07', count(station)=302)
Row(year=u'2013', month=u'05', count(station)=301)
Row(year=u'2013', month=u'07', count(station)=301)
Row(year=u'1962', month=u'08', count(station)=301)
Row(year=u'1996', month=u'10', count(station)=301)
Row(year=u'1964', month=u'07', count(station)=301)
Row(year=u'2007', month=u'05', count(station)=300)
Row(year=u'2013', month=u'08', count(station)=300)
Row(year=u'2007', month=u'09', count(station)=300)
Row(year=u'1962', month=u'07', count(station)=300)
Row(year=u'1962', month=u'09', count(station)=300)
```

```
Row(year=u'1964', month=u'08', count(station)=300)
Row(year=u'1994', month=u'08', count(station)=300)
Row(year=u'1966', month=u'04', count(station)=300)
Row(year=u'2013', month=u'09', count(station)=299)
Row(year=u'1994', month=u'05', count(station)=299)
Row(year=u'1994', month=u'09', count(station)=299)
Row(year=u'2014', month=u'06', count(station)=298)
Row(year=u'1986', month=u'10', count(station)=298)
Row(year=u'2005', month=u'10', count(station)=298)
Row(year=u'2014', month=u'07', count(station)=297)
Row(year=u'1962', month=u'06', count(station)=297)
Row(year=u'1964', month=u'09', count(station)=297)
Row(year=u'2008', month=u'04', count(station)=296)
Row(year=u'2014', month=u'05', count(station)=296)
Row(year=u'2014', month=u'08', count(station)=296)
Row(year=u'2014', month=u'09', count(station)=296)
Row(year=u'1995', month=u'06', count(station)=294)
Row(year=u'1990', month=u'10', count(station)=294)
Row(year=u'1986', month=u'07', count(station)=294)
Row(year=u'1976', month=u'04', count(station)=294)
Row(year=u'1993', month=u'08', count(station)=293)
Row(year=u'1995', month=u'05', count(station)=292)
Row(year=u'1962', month=u'05', count(station)=292)
Row(year=u'2003', month=u'04', count(station)=292)
Row(year=u'1993', month=u'05', count(station)=292)
Row(year=u'1995', month=u'07', count(station)=291)
Row(year=u'1969', month=u'04', count(station)=291)
Row(year=u'1993', month=u'07', count(station)=291)
Row(year=u'2002', month=u'04', count(station)=290)
Row(year=u'1991', month=u'05', count(station)=290)
Row(year=u'1993', month=u'06', count(station)=289)
Row(year=u'2011', month=u'04', count(station)=289)
Row(year=u'1964', month=u'10', count(station)=288)
Row(year=u'1963', month=u'10', count(station)=286)
Row(year=u'1988', month=u'10', count(station)=285)
Row(year=u'1963', month=u'04', count(station)=283)
Row(year=u'1993', month=u'09', count(station)=280)
Row(year=u'2001', month=u'10', count(station)=279)
Row(year=u'1990', month=u'04', count(station)=279)
Row(year=u'1967', month=u'04', count(station)=279)
Row(year=u'2011', month=u'10', count(station)=279)
Row(year=u'1993', month=u'04', count(station)=278)
Row(year=u'2010', month=u'10', count(station)=277)
Row(year=u'1999', month=u'04', count(station)=277)
Row(year=u'2006', month=u'10', count(station)=276)
Row(year=u'1973', month=u'03', count(station)=274)
Row(year=u'2009', month=u'04', count(station)=274)
Row(year=u'1979', month=u'10', count(station)=272)
Row(year=u'1998', month=u'04', count(station)=271)
Row(year=u'2000', month=u'04', count(station)=271)
Row(year=u'2013', month=u'10', count(station)=270)
Row(year=u'1962', month=u'04', count(station)=270)
Row(year=u'2014', month=u'10', count(station)=270)
```

```
Row(year=u'1961', month=u'08', count(station)=269)
Row(year=u'1961', month=u'09', count(station)=269)
Row(year=u'1991', month=u'10', count(station)=269)
Row(year=u'1974', month=u'10', count(station)=269)
Row(year=u'1975', month=u'04', count(station)=269)
Row(year=u'1994', month=u'04', count(station)=269)
Row(year=u'1976', month=u'10', count(station)=269)
Row(year=u'1989', month=u'04', count(station)=268)
Row(year=u'1961', month=u'05', count(station)=268)
Row(year=u'1961', month=u'06', count(station)=268)
Row(year=u'1961', month=u'10', count(station)=268)
Row(year=u'1992', month=u'10', count(station)=268)
Row(year=u'2004', month=u'04', count(station)=267)
Row(year=u'1961', month=u'07', count(station)=266)
Row(year=u'2007', month=u'10', count(station)=266)
Row(year=u'1966', month=u'10', count(station)=265)
Row(year=u'2007', month=u'04', count(station)=263)
Row(year=u'1996', month=u'04', count(station)=263)
Row(year=u'1991', month=u'04', count(station)=263)
Row(year=u'1964', month=u'04', count(station)=263)
Row(year=u'1980', month=u'04', count(station)=262)
Row(year=u'1987', month=u'04', count(station)=261)
Row(year=u'1986', month=u'04', count(station)=260)
Row(year=u'1994', month=u'10', count(station)=257)
Row(year=u'2004', month=u'10', count(station)=255)
Row(year=u'2005', month=u'04', count(station)=255)
Row(year=u'2014', month=u'04', count(station)=254)
Row(year=u'1980', month=u'10', count(station)=249)
Row(year=u'1981', month=u'04', count(station)=248)
Row(year=u'1983', month=u'10', count(station)=246)
Row(year=u'1982', month=u'04', count(station)=246)
Row(year=u'1968', month=u'10', count(station)=245)
Row(year=u'1972', month=u'04', count(station)=245)
Row(year=u'1961', month=u'04', count(station)=244)
Row(year=u'1973', month=u'04', count(station)=243)
Row(year=u'1999', month=u'10', count(station)=243)
Row(year=u'1971', month=u'04', count(station)=242)
Row(year=u'2012', month=u'10', count(station)=242)
Row(year=u'1978', month=u'04', count(station)=241)
Row(year=u'1971', month=u'11', count(station)=241)
Row(year=u'1989', month=u'10', count(station)=239)
Row(year=u'1977', month=u'04', count(station)=238)
Row(year=u'1983', month=u'04', count(station)=235)
Row(year=u'2001', month=u'04', count(station)=230)
Row(year=u'1978', month=u'11', count(station)=229)
Row(year=u'1997', month=u'10', count(station)=229)
Row(year=u'2006', month=u'04', count(station)=229)
Row(year=u'1998', month=u'10', count(station)=228)
Row(year=u'1979', month=u'04', count(station)=227)
Row(year=u'2008', month=u'10', count(station)=226)
Row(year=u'1999', month=u'11', count(station)=225)
Row(year=u'1974', month=u'03', count(station)=222)
Row(year=u'2010', month=u'04', count(station)=222)
```

```
Row(year=u'1988', month=u'04', count(station)=220)
Row(year=u'2012', month=u'04', count(station)=215)
Row(year=u'2013', month=u'04', count(station)=208)
Row(year=u'1993', month=u'10', count(station)=206)
Row(year=u'2003', month=u'10', count(station)=203)
Row(year=u'1968', month=u'03', count(station)=202)
Row(year=u'2007', month=u'03', count(station)=201)
Row(year=u'1961', month=u'03', count(station)=197)
Row(year=u'2011', month=u'11', count(station)=197)
Row(year=u'1982', month=u'03', count(station)=196)
Row(year=u'2005', month=u'11', count(station)=195)
Row(year=u'1977', month=u'11', count(station)=193)
Row(year=u'1990', month=u'03', count(station)=193)
Row(year=u'1997', month=u'04', count(station)=190)
Row(year=u'1995', month=u'04', count(station)=188)
Row(year=u'2002', month=u'10', count(station)=185)
Row(year=u'1992', month=u'04', count(station)=181)
Row(year=u'1982', month=u'11', count(station)=180)
Row(year=u'1984', month=u'11', count(station)=179)
Row(year=u'1996', month=u'11', count(station)=179)
Row(year=u'1965', month=u'03', count(station)=179)
Row(year=u'1972', month=u'11', count(station)=176)
Row(year=u'1972', month=u'03', count(station)=175)
Row(year=u'2012', month=u'03', count(station)=172)
Row(year=u'2014', month=u'03', count(station)=169)
Row(year=u'2009', month=u'10', count(station)=164)
Row(year=u'1975', month=u'11', count(station)=163)
Row(year=u'1991', month=u'03', count(station)=162)
Row(year=u'1983', month=u'11', count(station)=160)
Row(year=u'1967', month=u'03', count(station)=160)
Row(year=u'2014', month=u'11', count(station)=158)
Row(year=u'2000', month=u'03', count(station)=157)
Row(year=u'1997', month=u'03', count(station)=156)
Row(year=u'2002', month=u'03', count(station)=154)
Row(year=u'2001', month=u'11', count(station)=150)
Row(year=u'2007', month=u'11', count(station)=149)
Row(year=u'1990', month=u'02', count(station)=148)
Row(year=u'1999', month=u'03', count(station)=148)
Row(year=u'2006', month=u'11', count(station)=145)
Row(year=u'1998', month=u'03', count(station)=142)
Row(year=u'1961', month=u'11', count(station)=140)
Row(year=u'2003', month=u'03', count(station)=140)
Row(year=u'2004', month=u'03', count(station)=139)
Row(year=u'1986', month=u'11', count(station)=138)
Row(year=u'1969', month=u'11', count(station)=135)
Row(year=u'1985', month=u'04', count(station)=134)
Row(year=u'1989', month=u'03', count(station)=133)
Row(year=u'1970', month=u'04', count(station)=131)
Row(year=u'1957', month=u'07', count(station)=130)
Row(year=u'1957', month=u'08', count(station)=130)
Row(year=u'1958', month=u'06', count(station)=130)
Row(year=u'1958', month=u'08', count(station)=130)
Row(year=u'1957', month=u'09', count(station)=129)
```

```
Row(year=u'1958', month=u'07', count(station)=129)
Row(year=u'1958', month=u'09', count(station)=129)
Row(year=u'1989', month=u'11', count(station)=128)
Row(year=u'1957', month=u'06', count(station)=128)
Row(year=u'1959', month=u'06', count(station)=128)
Row(year=u'1956', month=u'06', count(station)=127)
Row(year=u'1956', month=u'07', count(station)=127)
Row(year=u'1956', month=u'08', count(station)=127)
Row(year=u'1956', month=u'09', count(station)=127)
Row(year=u'1958', month=u'05', count(station)=127)
Row(year=u'2005', month=u'03', count(station)=127)
Row(year=u'1959', month=u'05', count(station)=127)
Row(year=u'1959', month=u'09', count(station)=127)
Row(year=u'1960', month=u'05', count(station)=127)
Row(year=u'1960', month=u'06', count(station)=127)
Row(year=u'1960', month=u'08', count(station)=127)
Row(year=u'1960', month=u'09', count(station)=127)
Row(year=u'1955', month=u'09', count(station)=126)
Row(year=u'1957', month=u'05', count(station)=126)
Row(year=u'1959', month=u'07', count(station)=126)
Row(year=u'1960', month=u'07', count(station)=126)
Row(year=u'1959', month=u'10', count(station)=126)
Row(year=u'2006', month=u'12', count(station)=125)
Row(year=u'1955', month=u'06', count(station)=125)
Row(year=u'1956', month=u'05', count(station)=125)
Row(year=u'1959', month=u'08', count(station)=125)
Row(year=u'1955', month=u'07', count(station)=124)
Row(year=u'1955', month=u'08', count(station)=124)
Row(year=u'1958', month=u'10', count(station)=124)
Row(year=u'1981', month=u'03', count(station)=123)
Row(year=u'2004', month=u'11', count(station)=122)
Row(year=u'1955', month=u'05', count(station)=119)
Row(year=u'1953', month=u'06', count(station)=119)
Row(year=u'1953', month=u'08', count(station)=119)
Row(year=u'1954', month=u'05', count(station)=119)
Row(year=u'1954', month=u'06', count(station)=119)
Row(year=u'1954', month=u'07', count(station)=119)
Row(year=u'1954', month=u'08', count(station)=119)
Row(year=u'1954', month=u'09', count(station)=119)
Row(year=u'1953', month=u'05', count(station)=118)
Row(year=u'1953', month=u'07', count(station)=118)
Row(year=u'1955', month=u'10', count(station)=117)
Row(year=u'1953', month=u'09', count(station)=117)
Row(year=u'1973', month=u'11', count(station)=116)
Row(year=u'1952', month=u'07', count(station)=115)
Row(year=u'1952', month=u'08', count(station)=115)
Row(year=u'1959', month=u'04', count(station)=115)
Row(year=u'2013', month=u'11', count(station)=114)
Row(year=u'1952', month=u'05', count(station)=114)
Row(year=u'1952', month=u'06', count(station)=114)
Row(year=u'1952', month=u'09', count(station)=114)
Row(year=u'2009', month=u'11', count(station)=114)
Row(year=u'1953', month=u'10', count(station)=114)
```

```
Row(year=u'1951', month=u'10', count(station)=113)
Row(year=u'1951', month=u'08', count(station)=112)
Row(year=u'1951', month=u'09', count(station)=112)
Row(year=u'1951', month=u'07', count(station)=111)
Row(year=u'1951', month=u'06', count(station)=110)
Row(year=u'1957', month=u'04', count(station)=107)
Row(year=u'1952', month=u'04', count(station)=107)
Row(year=u'1994', month=u'11', count(station)=107)
Row(year=u'2008', month=u'11', count(station)=106)
Row(year=u'2011', month=u'03', count(station)=104)
Row(year=u'1953', month=u'04', count(station)=104)
Row(year=u'1960', month=u'10', count(station)=103)
Row(year=u'1956', month=u'10', count(station)=103)
Row(year=u'1978', month=u'03', count(station)=99)
Row(year=u'1977', month=u'03', count(station)=99)
Row(year=u'1951', month=u'05', count(station)=98)
Row(year=u'1960', month=u'04', count(station)=98)
Row(year=u'1993', month=u'03', count(station)=96)
Row(year=u'1957', month=u'10', count(station)=93)
Row(year=u'1954', month=u'10', count(station)=91)
Row(year=u'1994', month=u'03', count(station)=89)
Row(year=u'1951', month=u'04', count(station)=88)
Row(year=u'2000', month=u'11', count(station)=87)
Row(year=u'1968', month=u'11', count(station)=85)
Row(year=u'1956', month=u'04', count(station)=84)
Row(year=u'1955', month=u'04', count(station)=81)
Row(year=u'1958', month=u'04', count(station)=81)
Row(year=u'2003', month=u'11', count(station)=79)
Row(year=u'1953', month=u'03', count(station)=77)
Row(year=u'2008', month=u'03', count(station)=73)
Row(year=u'1967', month=u'11', count(station)=73)
Row(year=u'1992', month=u'03', count(station)=73)
Row(year=u'1966', month=u'11', count(station)=70)
Row(year=u'1957', month=u'03', count(station)=69)
Row(year=u'1997', month=u'11', count(station)=69)
Row(year=u'1965', month=u'11', count(station)=68)
Row(year=u'2010', month=u'03', count(station)=65)
Row(year=u'1954', month=u'04', count(station)=65)
Row(year=u'1962', month=u'11', count(station)=63)
Row(year=u'1959', month=u'03', count(station)=63)
Row(year=u'1952', month=u'10', count(station)=62)
Row(year=u'1956', month=u'03', count(station)=61)
Row(year=u'2011', month=u'12', count(station)=61)
Row(year=u'2012', month=u'02', count(station)=60)
Row(year=u'1995', month=u'03', count(station)=59)
Row(year=u'1998', month=u'02', count(station)=59)
Row(year=u'1953', month=u'12', count(station)=58)
Row(year=u'1990', month=u'11', count(station)=56)
Row(year=u'2012', month=u'11', count(station)=56)
Row(year=u'1980', month=u'11', count(station)=55)
Row(year=u'1986', month=u'12', count(station)=53)
Row(year=u'1950', month=u'09', count(station)=50)
Row(year=u'1950', month=u'07', count(station)=49)
```

```
Row(year=u'1950', month=u'08', count(station)=49)
Row(year=u'1957', month=u'11', count(station)=49)
Row(year=u'2010', month=u'11', count(station)=49)
Row(year=u'1961', month=u'02', count(station)=48)
Row(year=u'2007', month=u'01', count(station)=47)
Row(year=u'1950', month=u'06', count(station)=47)
Row(year=u'2009', month=u'03', count(station)=47)
Row(year=u'1950', month=u'05', count(station)=46)
Row(year=u'1950', month=u'10', count(station)=46)
Row(year=u'1981', month=u'11', count(station)=45)
Row(year=u'1976', month=u'03', count(station)=43)
Row(year=u'1987', month=u'11', count(station)=43)
Row(year=u'1953', month=u'11', count(station)=42)
Row(year=u'1970', month=u'11', count(station)=41)
Row(year=u'1989', month=u'02', count(station)=39)
Row(year=u'1992', month=u'11', count(station)=39)
Row(year=u'1950', month=u'04', count(station)=36)
Row(year=u'1979', month=u'12', count(station)=35)
Row(year=u'1964', month=u'03', count(station)=35)
Row(year=u'1954', month=u'03', count(station)=34)
Row(year=u'1976', month=u'11', count(station)=33)
Row(year=u'1966', month=u'03', count(station)=33)
Row(year=u'2005', month=u'01', count(station)=32)
Row(year=u'2002', month=u'02', count(station)=31)
Row(year=u'1991', month=u'11', count(station)=30)
Row(year=u'1975', month=u'03', count(station)=30)
Row(year=u'2000', month=u'12', count(station)=29)
Row(year=u'1971', month=u'12', count(station)=27)
Row(year=u'1950', month=u'03', count(station)=26)
Row(year=u'1983', month=u'01', count(station)=26)
Row(year=u'2008', month=u'02', count(station)=25)
Row(year=u'1995', month=u'11', count(station)=24)
Row(year=u'1987', month=u'03', count(station)=24)
Row(year=u'1989', month=u'01', count(station)=23)
Row(year=u'1992', month=u'01', count(station)=23)
Row(year=u'1964', month=u'12', count(station)=23)
Row(year=u'1955', month=u'11', count(station)=22)
Row(year=u'1951', month=u'11', count(station)=22)
Row(year=u'1954', month=u'11', count(station)=21)
Row(year=u'1979', month=u'11', count(station)=21)
Row(year=u'1960', month=u'11', count(station)=19)
Row(year=u'1974', month=u'11', count(station)=19)
Row(year=u'1963', month=u'11', count(station)=19)
Row(year=u'1964', month=u'11', count(station)=19)
Row(year=u'1959', month=u'11', count(station)=19)
Row(year=u'1960', month=u'03', count(station)=18)
Row(year=u'1976', month=u'02', count(station)=17)
Row(year=u'1959', month=u'02', count(station)=17)
Row(year=u'1983', month=u'03', count(station)=17)
Row(year=u'1994', month=u'12', count(station)=17)
Row(year=u'1985', month=u'11', count(station)=16)
Row(year=u'1971', month=u'01', count(station)=16)
Row(year=u'1977', month=u'12', count(station)=15)
```

```
Row(year=u'2014', month=u'02', count(station)=15)
Row(year=u'1993', month=u'02', count(station)=15)
Row(year=u'1973', month=u'01', count(station)=14)
Row(year=u'1986', month=u'03', count(station)=14)
Row(year=u'1975', month=u'02', count(station)=14)
Row(year=u'1971', month=u'03', count(station)=14)
Row(year=u'1958', month=u'11', count(station)=11)
Row(year=u'1979', month=u'03', count(station)=10)
Row(year=u'1975', month=u'01', count(station)=10)
Row(year=u'2001', month=u'02', count(station)=10)
Row(year=u'2013', month=u'03', count(station)=9)
Row(year=u'1972', month=u'12', count(station)=8)
Row(year=u'1989', month=u'12', count(station)=8)
Row(year=u'2013', month=u'12', count(station)=8)
Row(year=u'1956', month=u'12', count(station)=8)
Row(year=u'1988', month=u'03', count(station)=8)
Row(year=u'1961', month=u'12', count(station)=7)
Row(year=u'1955', month=u'12', count(station)=7)
Row(year=u'1953', month=u'02', count(station)=7)
Row(year=u'2013', month=u'02', count(station)=6)
Row(year=u'1984', month=u'12', count(station)=6)
Row(year=u'1967', month=u'12', count(station)=6)
Row(year=u'1991', month=u'12', count(station)=6)
Row(year=u'2004', month=u'02', count(station)=5)
Row(year=u'1951', month=u'12', count(station)=5)
Row(year=u'2006', month=u'03', count(station)=5)
Row(year=u'1983', month=u'12', count(station)=4)
Row(year=u'1973', month=u'02', count(station)=4)
Row(year=u'1974', month=u'02', count(station)=4)
Row(year=u'1956', month=u'11', count(station)=4)
Row(year=u'1992', month=u'02', count(station)=4)
Row(year=u'1998', month=u'01', count(station)=4)
Row(year=u'1957', month=u'12', count(station)=4)
Row(year=u'1958', month=u'03', count(station)=4)
Row(year=u'1958', month=u'12', count(station)=4)
Row(year=u'1954', month=u'12', count(station)=3)
Row(year=u'1955', month=u'02', count(station)=3)
Row(year=u'1990', month=u'01', count(station)=3)
Row(year=u'2007', month=u'12', count(station)=3)
Row(year=u'1996', month=u'03', count(station)=3)
Row(year=u'1980', month=u'03', count(station)=3)
Row(year=u'2000', month=u'02', count(station)=3)
Row(year=u'1959', month=u'01', count(station)=3)
Row(year=u'1954', month=u'01', count(station)=3)
Row(year=u'1960', month=u'12', count(station)=2)
Row(year=u'1955', month=u'01', count(station)=2)
Row(year=u'1990', month=u'12', count(station)=2)
Row(year=u'1956', month=u'01', count(station)=2)
Row(year=u'1956', month=u'02', count(station)=2)
Row(year=u'1950', month=u'11', count(station)=2)
Row(year=u'2003', month=u'02', count(station)=2)
Row(year=u'1997', month=u'02', count(station)=2)
Row(year=u'1957', month=u'01', count(station)=2)
```

```
Row(year=u'1957', month=u'02', count(station)=2)
Row(year=u'1993', month=u'12', count(station)=2)
Row(year=u'1988', month=u'01', count(station)=2)
Row(year=u'1995', month=u'02', count(station)=1)
Row(year=u'1955', month=u'03', count(station)=1)
Row(year=u'1984', month=u'01', count(station)=1)
Row(year=u'1984', month=u'03', count(station)=1)
Row(year=u'1962', month=u'02', count(station)=1)
Row(year=u'1962', month=u'03', count(station)=1)
Row(year=u'1991', month=u'01', count(station)=1)
Row(year=u'1979', month=u'01', count(station)=1)
Row(year=u'1950', month=u'12', count(station)=1)
Row(year=u'1951', month=u'02', count(station)=1)
Row(year=u'2014', month=u'12', count(station)=1)
Row(year=u'1963', month=u'03', count(station)=1)
Row(year=u'1952', month=u'03', count(station)=1)
Row(year=u'1970', month=u'03', count(station)=1)
Row(year=u'1958', month=u'01', count(station)=1)
Row(year=u'1958', month=u'02', count(station)=1)
Row(year=u'1993', month=u'01', count(station)=1)
Row(year=u'1952', month=u'11', count(station)=1)
Row(year=u'1975', month=u'12', count(station)=1)
Row(year=u'1993', month=u'11', count(station)=1)
Row(year=u'1987', month=u'12', count(station)=1)
Row(year=u'2005', month=u'12', count(station)=1)
Row(year=u'1960', month=u'01', count(station)=1)
Row(year=u'1960', month=u'02', count(station)=1)
Row(year=u'1954', month=u'02', count(station)=1)
Row(year=u'2001', month=u'03', count(station)=1)
Row(year=u'1988', month=u'11', count(station)=1)
```

SQL methods

```
Row(year=u'1972', month=u'10', count=378)
Row(year=u'1973', month=u'05', count=377)
Row(year=u'1973', month=u'06', count=377)
Row(year=u'1973', month=u'09', count=376)
Row(year=u'1972', month=u'08', count=376)
Row(year=u'1972', month=u'09', count=375)
Row(year=u'1972', month=u'06', count=375)
Row(year=u'1972', month=u'05', count=375)
Row(year=u'1971', month=u'08', count=375)
Row(year=u'1971', month=u'06', count=374)
Row(year=u'1971', month=u'09', count=374)
Row(year=u'1972', month=u'07', count=374)
Row(year=u'1971', month=u'05', count=373)
Row(year=u'1973', month=u'08', count=373)
Row(year=u'1974', month=u'06', count=372)
Row(year=u'1974', month=u'08', count=372)
Row(year=u'1974', month=u'09', count=370)
Row(year=u'1974', month=u'05', count=370)
Row(year=u'1973', month=u'07', count=370)
Row(year=u'1970', month=u'08', count=370)
```

```
Row(year=u'1971', month=u'07', count=370)
Row(year=u'1970', month=u'09', count=369)
Row(year=u'1970', month=u'06', count=369)
Row(year=u'1976', month=u'05', count=369)
Row(year=u'1975', month=u'09', count=369)
Row(year=u'1975', month=u'06', count=368)
Row(year=u'1976', month=u'06', count=368)
Row(year=u'1975', month=u'05', count=367)
Row(year=u'1975', month=u'08', count=367)
Row(year=u'1970', month=u'05', count=366)
Row(year=u'1976', month=u'09', count=365)
Row(year=u'1977', month=u'06', count=364)
Row(year=u'1967', month=u'05', count=363)
Row(year=u'1976', month=u'08', count=363)
Row(year=u'1974', month=u'07', count=362)
Row(year=u'1970', month=u'07', count=362)
Row(year=u'1967', month=u'09', count=361)
Row(year=u'1966', month=u'06', count=360)
Row(year=u'1966', month=u'09', count=360)
Row(year=u'1967', month=u'06', count=359)
Row(year=u'1966', month=u'08', count=359)
Row(year=u'1969', month=u'09', count=359)
Row(year=u'1978', month=u'09', count=358)
Row(year=u'1965', month=u'09', count=358)
Row(year=u'1967', month=u'08', count=358)
Row(year=u'1975', month=u'07', count=358)
Row(year=u'1968', month=u'06', count=357)
Row(year=u'1968', month=u'08', count=357)
Row(year=u'1969', month=u'08', count=357)
Row(year=u'1968', month=u'09', count=356)
Row(year=u'1976', month=u'07', count=356)
Row(year=u'1968', month=u'05', count=355)
Row(year=u'1965', month=u'06', count=355)
Row(year=u'1978', month=u'06', count=354)
Row(year=u'1965', month=u'08', count=354)
Row(year=u'1977', month=u'08', count=354)
Row(year=u'1966', month=u'05', count=354)
Row(year=u'1979', month=u'05', count=354)
Row(year=u'1968', month=u'07', count=353)
Row(year=u'1977', month=u'09', count=353)
Row(year=u'1969', month=u'06', count=352)
Row(year=u'1966', month=u'07', count=352)
Row(year=u'1978', month=u'05', count=352)
Row(year=u'1979', month=u'06', count=351)
Row(year=u'1979', month=u'09', count=351)
Row(year=u'1977', month=u'05', count=351)
Row(year=u'1967', month=u'07', count=351)
Row(year=u'1978', month=u'08', count=350)
Row(year=u'1977', month=u'07', count=350)
Row(year=u'1965', month=u'07', count=349)
Row(year=u'1973', month=u'10', count=349)
Row(year=u'1969', month=u'07', count=349)
Row(year=u'1971', month=u'10', count=347)
```

```
Row(year=u'1969', month=u'10', count=346)
Row(year=u'1979', month=u'07', count=345)
Row(year=u'1970', month=u'10', count=345)
Row(year=u'1996', month=u'06', count=345)
Row(year=u'1965', month=u'05', count=344)
Row(year=u'1974', month=u'04', count=344)
Row(year=u'1978', month=u'07', count=343)
Row(year=u'1996', month=u'05', count=342)
Row(year=u'1996', month=u'07', count=342)
Row(year=u'1996', month=u'08', count=341)
Row(year=u'1979', month=u'08', count=340)
Row(year=u'1997', month=u'09', count=340)
Row(year=u'1975', month=u'10', count=340)
Row(year=u'1996', month=u'09', count=340)
Row(year=u'1978', month=u'10', count=340)
Row(year=u'1982', month=u'06', count=339)
Row(year=u'1980', month=u'09', count=338)
Row(year=u'1997', month=u'06', count=338)
Row(year=u'1983', month=u'06', count=337)
Row(year=u'1997', month=u'08', count=337)
Row(year=u'1980', month=u'05', count=337)
Row(year=u'1981', month=u'05', count=337)
Row(year=u'1983', month=u'05', count=336)
Row(year=u'1969', month=u'05', count=335)
Row(year=u'1981', month=u'09', count=335)
Row(year=u'1965', month=u'10', count=335)
Row(year=u'1981', month=u'08', count=334)
Row(year=u'1982', month=u'09', count=334)
Row(year=u'1984', month=u'05', count=333)
Row(year=u'1997', month=u'07', count=333)
Row(year=u'1980', month=u'06', count=332)
Row(year=u'1983', month=u'09', count=332)
Row(year=u'1981', month=u'06', count=331)
Row(year=u'1999', month=u'06', count=330)
Row(year=u'1983', month=u'08', count=330)
Row(year=u'1980', month=u'08', count=330)
Row(year=u'1982', month=u'05', count=330)
Row(year=u'1999', month=u'07', count=329)
Row(year=u'1981', month=u'07', count=329)
Row(year=u'1999', month=u'09', count=328)
Row(year=u'1984', month=u'09', count=327)
Row(year=u'1985', month=u'09', count=327)
Row(year=u'1999', month=u'08', count=327)
Row(year=u'2002', month=u'06', count=326)
Row(year=u'1998', month=u'09', count=326)
Row(year=u'1982', month=u'08', count=326)
Row(year=u'1998', month=u'07', count=326)
Row(year=u'1998', month=u'06', count=326)
Row(year=u'1998', month=u'08', count=326)
Row(year=u'1985', month=u'05', count=325)
Row(year=u'2000', month=u'08', count=325)
Row(year=u'1981', month=u'10', count=325)
Row(year=u'1999', month=u'05', count=325)
```

```
Row(year=u'1967', month=u'10', count=324)
Row(year=u'1984', month=u'06', count=324)
Row(year=u'2001', month=u'07', count=324)
Row(year=u'2002', month=u'07', count=324)
Row(year=u'2002', month=u'05', count=324)
Row(year=u'1980', month=u'07', count=324)
Row(year=u'1985', month=u'06', count=324)
Row(year=u'2001', month=u'06', count=324)
Row(year=u'2002', month=u'09', count=323)
Row(year=u'1986', month=u'09', count=323)
Row(year=u'2000', month=u'05', count=323)
Row(year=u'2003', month=u'06', count=323)
Row(year=u'2001', month=u'08', count=323)
Row(year=u'1987', month=u'06', count=323)
Row(year=u'1987', month=u'09', count=323)
Row(year=u'1988', month=u'06', count=322)
Row(year=u'1968', month=u'04', count=322)
Row(year=u'2000', month=u'09', count=322)
Row(year=u'2001', month=u'09', count=322)
Row(year=u'2002', month=u'08', count=322)
Row(year=u'1998', month=u'05', count=322)
Row(year=u'1991', month=u'06', count=321)
Row(year=u'2003', month=u'05', count=321)
Row(year=u'1982', month=u'07', count=321)
Row(year=u'2004', month=u'09', count=321)
Row(year=u'2003', month=u'07', count=321)
Row(year=u'2004', month=u'05', count=321)
Row(year=u'2000', month=u'06', count=321)
Row(year=u'1984', month=u'10', count=321)
Row(year=u'2004', month=u'08', count=320)
Row(year=u'2003', month=u'08', count=320)
Row(year=u'2003', month=u'09', count=320)
Row(year=u'1987', month=u'08', count=320)
Row(year=u'1987', month=u'05', count=320)
Row(year=u'1988', month=u'05', count=320)
Row(year=u'2010', month=u'06', count=320)
Row(year=u'2000', month=u'07', count=320)
Row(year=u'2004', month=u'07', count=319)
Row(year=u'1997', month=u'05', count=319)
Row(year=u'2010', month=u'05', count=319)
Row(year=u'2011', month=u'07', count=319)
Row(year=u'2004', month=u'06', count=319)
Row(year=u'1987', month=u'07', count=319)
Row(year=u'1983', month=u'07', count=319)
Row(year=u'2010', month=u'07', count=318)
Row(year=u'2011', month=u'06', count=318)
Row(year=u'1995', month=u'08', count=318)
Row(year=u'1984', month=u'08', count=318)
Row(year=u'2010', month=u'08', count=318)
Row(year=u'1989', month=u'05', count=318)
Row(year=u'1985', month=u'10', count=317)
Row(year=u'1986', month=u'05', count=317)
Row(year=u'1987', month=u'10', count=317)
```

```
Row(year=u'2008', month=u'05', count=316)
Row(year=u'2008', month=u'06', count=316)
Row(year=u'2010', month=u'09', count=316)
Row(year=u'1990', month=u'08', count=316)
Row(year=u'1989', month=u'08', count=316)
Row(year=u'1988', month=u'08', count=316)
Row(year=u'2011', month=u'08', count=316)
Row(year=u'1989', month=u'09', count=316)
Row(year=u'1995', month=u'09', count=315)
Row(year=u'1982', month=u'10', count=315)
Row(year=u'1988', month=u'09', count=315)
Row(year=u'2000', month=u'10', count=315)
Row(year=u'2011', month=u'09', count=315)
Row(year=u'1989', month=u'06', count=315)
Row(year=u'2011', month=u'05', count=315)
Row(year=u'1991', month=u'08', count=315)
Row(year=u'2001', month=u'05', count=315)
Row(year=u'2008', month=u'07', count=314)
Row(year=u'1985', month=u'08', count=314)
Row(year=u'1991', month=u'07', count=314)
Row(year=u'1986', month=u'06', count=314)
Row(year=u'1984', month=u'04', count=313)
Row(year=u'1990', month=u'06', count=313)
Row(year=u'2008', month=u'08', count=313)
Row(year=u'1989', month=u'07', count=312)
Row(year=u'1990', month=u'09', count=312)
Row(year=u'1984', month=u'07', count=312)
Row(year=u'2009', month=u'07', count=312)
Row(year=u'2009', month=u'09', count=312)
Row(year=u'1992', month=u'05', count=311)
Row(year=u'2009', month=u'08', count=311)
Row(year=u'2005', month=u'06', count=311)
Row(year=u'1963', month=u'08', count=311)
Row(year=u'1991', month=u'09', count=311)
Row(year=u'1992', month=u'07', count=311)
Row(year=u'1963', month=u'05', count=310)
Row(year=u'1992', month=u'08', count=310)
Row(year=u'1992', month=u'06', count=310)
Row(year=u'2012', month=u'06', count=310)
Row(year=u'2009', month=u'06', count=310)
Row(year=u'1963', month=u'06', count=310)
Row(year=u'1963', month=u'09', count=310)
Row(year=u'2006', month=u'06', count=310)
Row(year=u'2012', month=u'07', count=310)
Row(year=u'2012', month=u'08', count=310)
Row(year=u'2006', month=u'07', count=310)
Row(year=u'2006', month=u'09', count=309)
Row(year=u'2006', month=u'08', count=309)
Row(year=u'2009', month=u'05', count=308)
Row(year=u'1963', month=u'07', count=308)
Row(year=u'2012', month=u'05', count=308)
Row(year=u'1990', month=u'07', count=308)
Row(year=u'2005', month=u'07', count=307)
```

```
Row(year=u'2006', month=u'05', count=307)
Row(year=u'1990', month=u'05', count=307)
Row(year=u'2005', month=u'09', count=307)
Row(year=u'2012', month=u'09', count=306)
Row(year=u'2005', month=u'08', count=306)
Row(year=u'1988', month=u'07', count=306)
Row(year=u'1962', month=u'10', count=305)
Row(year=u'1985', month=u'07', count=304)
Row(year=u'2007', month=u'07', count=304)
Row(year=u'1964', month=u'05', count=304)
Row(year=u'1965', month=u'04', count=304)
Row(year=u'1986', month=u'08', count=304)
Row(year=u'1992', month=u'09', count=304)
Row(year=u'1994', month=u'06', count=303)
Row(year=u'1964', month=u'06', count=303)
Row(year=u'2007', month=u'06', count=303)
Row(year=u'1977', month=u'10', count=303)
Row(year=u'1995', month=u'10', count=302)
Row(year=u'2008', month=u'09', count=302)
Row(year=u'2005', month=u'05', count=302)
Row(year=u'2007', month=u'08', count=302)
Row(year=u'2013', month=u'06', count=302)
Row(year=u'1994', month=u'07', count=302)
Row(year=u'2013', month=u'07', count=301)
Row(year=u'1964', month=u'07', count=301)
Row(year=u'2013', month=u'05', count=301)
Row(year=u'1962', month=u'08', count=301)
Row(year=u'1996', month=u'10', count=301)
Row(year=u'2013', month=u'08', count=300)
Row(year=u'1962', month=u'09', count=300)
Row(year=u'2007', month=u'09', count=300)
Row(year=u'1962', month=u'07', count=300)
Row(year=u'2007', month=u'05', count=300)
Row(year=u'1964', month=u'08', count=300)
Row(year=u'1966', month=u'04', count=300)
Row(year=u'1994', month=u'08', count=300)
Row(year=u'2013', month=u'09', count=299)
Row(year=u'1994', month=u'05', count=299)
Row(year=u'1994', month=u'09', count=299)
Row(year=u'2014', month=u'06', count=298)
Row(year=u'1986', month=u'10', count=298)
Row(year=u'2005', month=u'10', count=298)
Row(year=u'1962', month=u'06', count=297)
Row(year=u'2014', month=u'07', count=297)
Row(year=u'1964', month=u'09', count=297)
Row(year=u'2014', month=u'08', count=296)
Row(year=u'2014', month=u'09', count=296)
Row(year=u'2008', month=u'04', count=296)
Row(year=u'2014', month=u'05', count=296)
Row(year=u'1976', month=u'04', count=294)
Row(year=u'1986', month=u'07', count=294)
Row(year=u'1990', month=u'10', count=294)
Row(year=u'1995', month=u'06', count=294)
```

```
Row(year=u'1993', month=u'08', count=293)
Row(year=u'1962', month=u'05', count=292)
Row(year=u'1993', month=u'05', count=292)
Row(year=u'1995', month=u'05', count=292)
Row(year=u'2003', month=u'04', count=292)
Row(year=u'1969', month=u'04', count=291)
Row(year=u'1995', month=u'07', count=291)
Row(year=u'1993', month=u'07', count=291)
Row(year=u'2002', month=u'04', count=290)
Row(year=u'1991', month=u'05', count=290)
Row(year=u'2011', month=u'04', count=289)
Row(year=u'1993', month=u'06', count=289)
Row(year=u'1964', month=u'10', count=288)
Row(year=u'1963', month=u'10', count=286)
Row(year=u'1988', month=u'10', count=285)
Row(year=u'1963', month=u'04', count=283)
Row(year=u'1993', month=u'09', count=280)
Row(year=u'1990', month=u'04', count=279)
Row(year=u'2001', month=u'10', count=279)
Row(year=u'1967', month=u'04', count=279)
Row(year=u'2011', month=u'10', count=279)
Row(year=u'1993', month=u'04', count=278)
Row(year=u'1999', month=u'04', count=277)
Row(year=u'2010', month=u'10', count=277)
Row(year=u'2006', month=u'10', count=276)
Row(year=u'1973', month=u'03', count=274)
Row(year=u'2009', month=u'04', count=274)
Row(year=u'1979', month=u'10', count=272)
Row(year=u'1998', month=u'04', count=271)
Row(year=u'2000', month=u'04', count=271)
Row(year=u'1962', month=u'04', count=270)
Row(year=u'2014', month=u'10', count=270)
Row(year=u'2013', month=u'10', count=270)
Row(year=u'1961', month=u'09', count=269)
Row(year=u'1991', month=u'10', count=269)
Row(year=u'1975', month=u'04', count=269)
Row(year=u'1976', month=u'10', count=269)
Row(year=u'1974', month=u'10', count=269)
Row(year=u'1961', month=u'08', count=269)
Row(year=u'1994', month=u'04', count=269)
Row(year=u'1989', month=u'04', count=268)
Row(year=u'1961', month=u'06', count=268)
Row(year=u'1961', month=u'05', count=268)
Row(year=u'1992', month=u'10', count=268)
Row(year=u'1961', month=u'10', count=268)
Row(year=u'2004', month=u'04', count=267)
Row(year=u'1961', month=u'07', count=266)
Row(year=u'2007', month=u'10', count=266)
Row(year=u'1966', month=u'10', count=265)
Row(year=u'1964', month=u'04', count=263)
Row(year=u'2007', month=u'04', count=263)
Row(year=u'1996', month=u'04', count=263)
Row(year=u'1991', month=u'04', count=263)
```

```
Row(year=u'1980', month=u'04', count=262)
Row(year=u'1987', month=u'04', count=261)
Row(year=u'1986', month=u'04', count=260)
Row(year=u'1994', month=u'10', count=257)
Row(year=u'2005', month=u'04', count=255)
Row(year=u'2004', month=u'10', count=255)
Row(year=u'2014', month=u'04', count=254)
Row(year=u'1980', month=u'10', count=249)
Row(year=u'1981', month=u'04', count=248)
Row(year=u'1982', month=u'04', count=246)
Row(year=u'1983', month=u'10', count=246)
Row(year=u'1968', month=u'10', count=245)
Row(year=u'1972', month=u'04', count=245)
Row(year=u'1961', month=u'04', count=244)
Row(year=u'1973', month=u'04', count=243)
Row(year=u'1999', month=u'10', count=243)
Row(year=u'2012', month=u'10', count=242)
Row(year=u'1971', month=u'04', count=242)
Row(year=u'1978', month=u'04', count=241)
Row(year=u'1971', month=u'11', count=241)
Row(year=u'1989', month=u'10', count=239)
Row(year=u'1977', month=u'04', count=238)
Row(year=u'1983', month=u'04', count=235)
Row(year=u'2001', month=u'04', count=230)
Row(year=u'1978', month=u'11', count=229)
Row(year=u'2006', month=u'04', count=229)
Row(year=u'1997', month=u'10', count=229)
Row(year=u'1998', month=u'10', count=228)
Row(year=u'1979', month=u'04', count=227)
Row(year=u'2008', month=u'10', count=226)
Row(year=u'1999', month=u'11', count=225)
Row(year=u'2010', month=u'04', count=222)
Row(year=u'1974', month=u'03', count=222)
Row(year=u'1988', month=u'04', count=220)
Row(year=u'2012', month=u'04', count=215)
Row(year=u'2013', month=u'04', count=208)
Row(year=u'1993', month=u'10', count=206)
Row(year=u'2003', month=u'10', count=203)
Row(year=u'1968', month=u'03', count=202)
Row(year=u'2007', month=u'03', count=201)
Row(year=u'2011', month=u'11', count=197)
Row(year=u'1961', month=u'03', count=197)
Row(year=u'1982', month=u'03', count=196)
Row(year=u'2005', month=u'11', count=195)
Row(year=u'1977', month=u'11', count=193)
Row(year=u'1990', month=u'03', count=193)
Row(year=u'1997', month=u'04', count=190)
Row(year=u'1995', month=u'04', count=188)
Row(year=u'2002', month=u'10', count=185)
Row(year=u'1992', month=u'04', count=181)
Row(year=u'1982', month=u'11', count=180)
Row(year=u'1996', month=u'11', count=179)
Row(year=u'1984', month=u'11', count=179)
```

```
Row(year=u'1965', month=u'03', count=179)
Row(year=u'1972', month=u'11', count=176)
Row(year=u'1972', month=u'03', count=175)
Row(year=u'2012', month=u'03', count=172)
Row(year=u'2014', month=u'03', count=169)
Row(year=u'2009', month=u'10', count=164)
Row(year=u'1975', month=u'11', count=163)
Row(year=u'1991', month=u'03', count=162)
Row(year=u'1967', month=u'03', count=160)
Row(year=u'1983', month=u'11', count=160)
Row(year=u'2014', month=u'11', count=158)
Row(year=u'2000', month=u'03', count=157)
Row(year=u'1997', month=u'03', count=156)
Row(year=u'2002', month=u'03', count=154)
Row(year=u'2001', month=u'11', count=150)
Row(year=u'2007', month=u'11', count=149)
Row(year=u'1990', month=u'02', count=148)
Row(year=u'1999', month=u'03', count=148)
Row(year=u'2006', month=u'11', count=145)
Row(year=u'1998', month=u'03', count=142)
Row(year=u'1961', month=u'11', count=140)
Row(year=u'2003', month=u'03', count=140)
Row(year=u'2004', month=u'03', count=139)
Row(year=u'1986', month=u'11', count=138)
Row(year=u'1969', month=u'11', count=135)
Row(year=u'1985', month=u'04', count=134)
Row(year=u'1989', month=u'03', count=133)
Row(year=u'1970', month=u'04', count=131)
Row(year=u'1957', month=u'07', count=130)
Row(year=u'1958', month=u'08', count=130)
Row(year=u'1958', month=u'06', count=130)
Row(year=u'1957', month=u'08', count=130)
Row(year=u'1957', month=u'09', count=129)
Row(year=u'1958', month=u'07', count=129)
Row(year=u'1958', month=u'09', count=129)
Row(year=u'1989', month=u'11', count=128)
Row(year=u'1959', month=u'06', count=128)
Row(year=u'1957', month=u'06', count=128)
Row(year=u'1956', month=u'08', count=127)
Row(year=u'1956', month=u'09', count=127)
Row(year=u'2005', month=u'03', count=127)
Row(year=u'1960', month=u'08', count=127)
Row(year=u'1959', month=u'05', count=127)
Row(year=u'1960', month=u'09', count=127)
Row(year=u'1956', month=u'06', count=127)
Row(year=u'1959', month=u'09', count=127)
Row(year=u'1956', month=u'07', count=127)
Row(year=u'1958', month=u'05', count=127)
Row(year=u'1960', month=u'05', count=127)
Row(year=u'1960', month=u'06', count=127)
Row(year=u'1957', month=u'05', count=126)
Row(year=u'1959', month=u'07', count=126)
Row(year=u'1959', month=u'10', count=126)
```

```
Row(year=u'1955', month=u'09', count=126)
Row(year=u'1960', month=u'07', count=126)
Row(year=u'2006', month=u'12', count=125)
Row(year=u'1956', month=u'05', count=125)
Row(year=u'1959', month=u'08', count=125)
Row(year=u'1955', month=u'06', count=125)
Row(year=u'1958', month=u'10', count=124)
Row(year=u'1955', month=u'07', count=124)
Row(year=u'1955', month=u'08', count=124)
Row(year=u'1981', month=u'03', count=123)
Row(year=u'2004', month=u'11', count=122)
Row(year=u'1954', month=u'06', count=119)
Row(year=u'1954', month=u'05', count=119)
Row(year=u'1953', month=u'06', count=119)
Row(year=u'1954', month=u'09', count=119)
Row(year=u'1955', month=u'05', count=119)
Row(year=u'1954', month=u'08', count=119)
Row(year=u'1953', month=u'08', count=119)
Row(year=u'1954', month=u'07', count=119)
Row(year=u'1953', month=u'07', count=118)
Row(year=u'1953', month=u'05', count=118)
Row(year=u'1955', month=u'10', count=117)
Row(year=u'1953', month=u'09', count=117)
Row(year=u'1973', month=u'11', count=116)
Row(year=u'1952', month=u'08', count=115)
Row(year=u'1959', month=u'04', count=115)
Row(year=u'1952', month=u'07', count=115)
Row(year=u'2013', month=u'11', count=114)
Row(year=u'1952', month=u'06', count=114)
Row(year=u'1953', month=u'10', count=114)
Row(year=u'1952', month=u'05', count=114)
Row(year=u'1952', month=u'09', count=114)
Row(year=u'2009', month=u'11', count=114)
Row(year=u'1951', month=u'10', count=113)
Row(year=u'1951', month=u'09', count=112)
Row(year=u'1951', month=u'08', count=112)
Row(year=u'1951', month=u'07', count=111)
Row(year=u'1951', month=u'06', count=110)
Row(year=u'1952', month=u'04', count=107)
Row(year=u'1957', month=u'04', count=107)
Row(year=u'1994', month=u'11', count=107)
Row(year=u'2008', month=u'11', count=106)
Row(year=u'2011', month=u'03', count=104)
Row(year=u'1953', month=u'04', count=104)
Row(year=u'1956', month=u'10', count=103)
Row(year=u'1960', month=u'10', count=103)
Row(year=u'1978', month=u'03', count=99)
Row(year=u'1977', month=u'03', count=99)
Row(year=u'1951', month=u'05', count=98)
Row(year=u'1960', month=u'04', count=98)
Row(year=u'1993', month=u'03', count=96)
Row(year=u'1957', month=u'10', count=93)
Row(year=u'1954', month=u'10', count=91)
```

```
Row(year=u'1994', month=u'03', count=89)
Row(year=u'1951', month=u'04', count=88)
Row(year=u'2000', month=u'11', count=87)
Row(year=u'1968', month=u'11', count=85)
Row(year=u'1956', month=u'04', count=84)
Row(year=u'1955', month=u'04', count=81)
Row(year=u'1958', month=u'04', count=81)
Row(year=u'2003', month=u'11', count=79)
Row(year=u'1953', month=u'03', count=77)
Row(year=u'1967', month=u'11', count=73)
Row(year=u'1992', month=u'03', count=73)
Row(year=u'2008', month=u'03', count=73)
Row(year=u'1966', month=u'11', count=70)
Row(year=u'1957', month=u'03', count=69)
Row(year=u'1997', month=u'11', count=69)
Row(year=u'1965', month=u'11', count=68)
Row(year=u'2010', month=u'03', count=65)
Row(year=u'1954', month=u'04', count=65)
Row(year=u'1959', month=u'03', count=63)
Row(year=u'1962', month=u'11', count=63)
Row(year=u'1952', month=u'10', count=62)
Row(year=u'2011', month=u'12', count=61)
Row(year=u'1956', month=u'03', count=61)
Row(year=u'2012', month=u'02', count=60)
Row(year=u'1995', month=u'03', count=59)
Row(year=u'1998', month=u'02', count=59)
Row(year=u'1953', month=u'12', count=58)
Row(year=u'1990', month=u'11', count=56)
Row(year=u'2012', month=u'11', count=56)
Row(year=u'1980', month=u'11', count=55)
Row(year=u'1986', month=u'12', count=53)
Row(year=u'1950', month=u'09', count=50)
Row(year=u'1950', month=u'08', count=49)
Row(year=u'1950', month=u'07', count=49)
Row(year=u'2010', month=u'11', count=49)
Row(year=u'1957', month=u'11', count=49)
Row(year=u'1961', month=u'02', count=48)
Row(year=u'1950', month=u'06', count=47)
Row(year=u'2009', month=u'03', count=47)
Row(year=u'2007', month=u'01', count=47)
Row(year=u'1950', month=u'05', count=46)
Row(year=u'1950', month=u'10', count=46)
Row(year=u'1981', month=u'11', count=45)
Row(year=u'1976', month=u'03', count=43)
Row(year=u'1987', month=u'11', count=43)
Row(year=u'1953', month=u'11', count=42)
Row(year=u'1970', month=u'11', count=41)
Row(year=u'1989', month=u'02', count=39)
Row(year=u'1992', month=u'11', count=39)
Row(year=u'1950', month=u'04', count=36)
Row(year=u'1964', month=u'03', count=35)
Row(year=u'1979', month=u'12', count=35)
Row(year=u'1954', month=u'03', count=34)
```

```
Row(year=u'1976', month=u'11', count=33)
Row(year=u'1966', month=u'03', count=33)
Row(year=u'2005', month=u'01', count=32)
Row(year=u'2002', month=u'02', count=31)
Row(year=u'1975', month=u'03', count=30)
Row(year=u'1991', month=u'11', count=30)
Row(year=u'2000', month=u'12', count=29)
Row(year=u'1971', month=u'12', count=27)
Row(year=u'1950', month=u'03', count=26)
Row(year=u'1983', month=u'01', count=26)
Row(year=u'2008', month=u'02', count=25)
Row(year=u'1995', month=u'11', count=24)
Row(year=u'1987', month=u'03', count=24)
Row(year=u'1964', month=u'12', count=23)
Row(year=u'1992', month=u'01', count=23)
Row(year=u'1989', month=u'01', count=23)
Row(year=u'1951', month=u'11', count=22)
Row(year=u'1955', month=u'11', count=22)
Row(year=u'1954', month=u'11', count=21)
Row(year=u'1979', month=u'11', count=21)
Row(year=u'1963', month=u'11', count=19)
Row(year=u'1960', month=u'11', count=19)
Row(year=u'1974', month=u'11', count=19)
Row(year=u'1959', month=u'11', count=19)
Row(year=u'1964', month=u'11', count=19)
Row(year=u'1960', month=u'03', count=18)
Row(year=u'1983', month=u'03', count=17)
Row(year=u'1976', month=u'02', count=17)
Row(year=u'1994', month=u'12', count=17)
Row(year=u'1959', month=u'02', count=17)
Row(year=u'1985', month=u'11', count=16)
Row(year=u'1971', month=u'01', count=16)
Row(year=u'2014', month=u'02', count=15)
Row(year=u'1977', month=u'12', count=15)
Row(year=u'1993', month=u'02', count=15)
Row(year=u'1986', month=u'03', count=14)
Row(year=u'1973', month=u'01', count=14)
Row(year=u'1971', month=u'03', count=14)
Row(year=u'1975', month=u'02', count=14)
Row(year=u'1958', month=u'11', count=11)
Row(year=u'1975', month=u'01', count=10)
Row(year=u'2001', month=u'02', count=10)
Row(year=u'1979', month=u'03', count=10)
Row(year=u'2013', month=u'03', count=9)
Row(year=u'2013', month=u'12', count=8)
Row(year=u'1989', month=u'12', count=8)
Row(year=u'1956', month=u'12', count=8)
Row(year=u'1988', month=u'03', count=8)
Row(year=u'1972', month=u'12', count=8)
Row(year=u'1955', month=u'12', count=7)
Row(year=u'1961', month=u'12', count=7)
Row(year=u'1953', month=u'02', count=7)
Row(year=u'1984', month=u'12', count=6)
```

```
Row(year=u'1991', month=u'12', count=6)
Row(year=u'2013', month=u'02', count=6)
Row(year=u'1967', month=u'12', count=6)
Row(year=u'2006', month=u'03', count=5)
Row(year=u'1951', month=u'12', count=5)
Row(year=u'2004', month=u'02', count=5)
Row(year=u'1998', month=u'01', count=4)
Row(year=u'1958', month=u'03', count=4)
Row(year=u'1992', month=u'02', count=4)
Row(year=u'1958', month=u'12', count=4)
Row(year=u'1957', month=u'12', count=4)
Row(year=u'1973', month=u'02', count=4)
Row(year=u'1974', month=u'02', count=4)
Row(year=u'1983', month=u'12', count=4)
Row(year=u'1956', month=u'11', count=4)
Row(year=u'1954', month=u'12', count=3)
Row(year=u'1990', month=u'01', count=3)
Row(year=u'1954', month=u'01', count=3)
Row(year=u'2007', month=u'12', count=3)
Row(year=u'1955', month=u'02', count=3)
Row(year=u'1980', month=u'03', count=3)
Row(year=u'1996', month=u'03', count=3)
Row(year=u'1959', month=u'01', count=3)
Row(year=u'2000', month=u'02', count=3)
Row(year=u'1955', month=u'01', count=2)
Row(year=u'1957', month=u'02', count=2)
Row(year=u'1960', month=u'12', count=2)
Row(year=u'1956', month=u'01', count=2)
Row(year=u'1950', month=u'11', count=2)
Row(year=u'2003', month=u'02', count=2)
Row(year=u'1990', month=u'12', count=2)
Row(year=u'1997', month=u'02', count=2)
Row(year=u'1956', month=u'02', count=2)
Row(year=u'1957', month=u'01', count=2)
Row(year=u'1993', month=u'12', count=2)
Row(year=u'1988', month=u'01', count=2)
Row(year=u'1962', month=u'03', count=1)
Row(year=u'1975', month=u'12', count=1)
Row(year=u'1979', month=u'01', count=1)
Row(year=u'1984', month=u'01', count=1)
Row(year=u'1951', month=u'02', count=1)
Row(year=u'1963', month=u'03', count=1)
Row(year=u'1958', month=u'02', count=1)
Row(year=u'1993', month=u'01', count=1)
Row(year=u'1987', month=u'12', count=1)
Row(year=u'1954', month=u'02', count=1)
Row(year=u'1984', month=u'03', count=1)
Row(year=u'1955', month=u'03', count=1)
Row(year=u'1988', month=u'11', count=1)
Row(year=u'1950', month=u'12', count=1)
Row(year=u'2014', month=u'12', count=1)
Row(year=u'1952', month=u'03', count=1)
Row(year=u'2005', month=u'12', count=1)
```

```
Row(year=u'1960', month=u'02', count=1)
Row(year=u'1993', month=u'11', count=1)
Row(year=u'2001', month=u'03', count=1)
Row(year=u'1995', month=u'02', count=1)
Row(year=u'1962', month=u'02', count=1)
Row(year=u'1958', month=u'01', count=1)
Row(year=u'1970', month=u'03', count=1)
Row(year=u'1952', month=u'11', count=1)
Row(year=u'1991', month=u'01', count=1)
Row(year=u'1960', month=u'01', count=1)
```

3.Find the average monthly temperature for each available station in Sweden. Your result should include average temperature for each sation for each month in period of 1960-2014. Note that not every station has the readings for each month in this timeframe.

year, month, station, avgMonthlyTemperature ORDER BY avgMonthlyTempearture DESC

```python
from pyspark import SparkContext
from pyspark.sql import functions as F
from pyspark.sql import SQLContext, Row

sc = SparkContext(appName="BDA2_Q3")
sqlContext = SQLContext(sc)

###Read the data
data = sc.textFile("/user/x_saeju/data/temperature-readings.csv").cache()
data = data.map(lambda line: line.split(";"))
data = data.filter(lambda x:(int(x[1][0:4])>=1960 and int(x[1][0:4])<=2014))
data = data.map(lambda x:Row(station=x[0], date=x[1], year=x[1].split("-")[0],
                             month=x[1].split("-")[1], day=x[1].split("-")[2],
                             time=x[2], temp=float(x[3]), quality=x[4]))

q3data = sqlContext.createDataFrame(data)
q3data.registerTempTable("Q3_dataframe")


###Calculate avgMonthlyTemperature
####for each date
minTemp = q3data.groupBy("year", "month", "day", "station").agg(F.min("temp").alias("minT"))
maxTemp = q3data.groupBy("year", "month", "day", "station").agg(F.max("temp").alias("maxT"))
#The alias, like in SQL, allows you to distinguish where each column is coming from.
#Now we have both min and max Temperature values for each station, date(year-month-day)
AvgTemp = minTemp.join(maxTemp,["year", "month", "day", "station"]) #not choosing left/rightjoin to avo
#Now we have minTemp and maxTemp in same table corresponding to each(same) station and date
#create new column - AvgTemp for each date
AvgTemp = AvgTemp.withColumn("AvgdateTemp",(AvgTemp.minT+AvgTemp.maxT)/2)
AvgTemp = AvgTemp.select("year", "month", "station", "AvgdateTemp")
#What we need is avgMONTHLYtemp, so we need one more average step.
AvgMonTemp = AvgTemp.groupBy("year","month","station").agg(F.avg("AvgdateTemp").alias("AvgMonthTemp"))
                       .orderBy(["AvgMonthTemp"],ascending=[0])
AvgMonTemp = AvgMonTemp.rdd.repartition(1).sortBy(ascending=False, keyfunc=lambda x:x[3])
#sorted RDD with the help of sortBy()
AvgMonTemp.repartition(1).saveAsTextFile("Q3_AvgMonTemp")
```

Results(first 100 lines)

```
Row(year=u'2014', month=u'07', station=u'96000', AvgMonthTemp=26.3)
Row(year=u'1994', month=u'07', station=u'96550', AvgMonthTemp=23.07105263157895)
Row(year=u'1983', month=u'08', station=u'54550', AvgMonthTemp=23.0)
Row(year=u'1994', month=u'07', station=u'78140', AvgMonthTemp=22.970967741935482)
Row(year=u'1994', month=u'07', station=u'85280', AvgMonthTemp=22.872580645161293)
Row(year=u'1994', month=u'07', station=u'75120', AvgMonthTemp=22.85806451612903)
Row(year=u'1994', month=u'07', station=u'65450', AvgMonthTemp=22.85645161290323)
Row(year=u'1994', month=u'07', station=u'96000', AvgMonthTemp=22.808064516129033)
Row(year=u'1994', month=u'07', station=u'95160', AvgMonthTemp=22.764516129032256)
Row(year=u'1994', month=u'07', station=u'86200', AvgMonthTemp=22.71129032258064)
Row(year=u'2002', month=u'08', station=u'78140', AvgMonthTemp=22.700000000000006)
Row(year=u'1994', month=u'07', station=u'76000', AvgMonthTemp=22.698387096774198)
Row(year=u'1997', month=u'08', station=u'78140', AvgMonthTemp=22.666129032258066)
Row(year=u'1994', month=u'07', station=u'105260', AvgMonthTemp=22.659677419354843)
Row(year=u'1975', month=u'08', station=u'54550', AvgMonthTemp=22.642857142857142)
Row(year=u'2006', month=u'07', station=u'76530', AvgMonthTemp=22.598387096774193)
Row(year=u'1994', month=u'07', station=u'86330', AvgMonthTemp=22.548387096774185)
Row(year=u'2006', month=u'07', station=u'75120', AvgMonthTemp=22.527419354838706)
Row(year=u'1994', month=u'07', station=u'54300', AvgMonthTemp=22.46935483870968)
Row(year=u'2006', month=u'07', station=u'78140', AvgMonthTemp=22.458064516129035)
Row(year=u'2001', month=u'07', station=u'96550', AvgMonthTemp=22.408333333333335)
Row(year=u'2010', month=u'07', station=u'98180', AvgMonthTemp=22.379032258064512)
Row(year=u'2006', month=u'07', station=u'65450', AvgMonthTemp=22.377419354838707)
Row(year=u'1994', month=u'07', station=u'85210', AvgMonthTemp=22.3758064516129)
Row(year=u'2014', month=u'07', station=u'98180', AvgMonthTemp=22.36774193548387)
Row(year=u'1994', month=u'07', station=u'98180', AvgMonthTemp=22.367741935483867)
Row(year=u'2002', month=u'08', station=u'98180', AvgMonthTemp=22.366129032258062)
Row(year=u'1994', month=u'07', station=u'92100', AvgMonthTemp=22.31774193548387)
Row(year=u'1994', month=u'07', station=u'86470', AvgMonthTemp=22.308064516129033)
Row(year=u'1994', month=u'07', station=u'83230', AvgMonthTemp=22.272580645161288)
Row(year=u'1994', month=u'07', station=u'64290', AvgMonthTemp=22.25967741935484)
Row(year=u'1994', month=u'07', station=u'97490', AvgMonthTemp=22.258064516129036)
Row(year=u'1994', month=u'07', station=u'94180', AvgMonthTemp=22.253225806451606)
Row(year=u'1972', month=u'07', station=u'173960', AvgMonthTemp=22.244999999999997)
Row(year=u'1994', month=u'07', station=u'74080', AvgMonthTemp=22.241935483870968)
Row(year=u'2006', month=u'07', station=u'54300', AvgMonthTemp=22.23709677419355)
Row(year=u'2002', month=u'08', station=u'98210', AvgMonthTemp=22.23548387096774)
Row(year=u'1994', month=u'07', station=u'106070', AvgMonthTemp=22.23225806451613)
Row(year=u'1994', month=u'07', station=u'75100', AvgMonthTemp=22.229032258064514)
Row(year=u'1994', month=u'07', station=u'53440', AvgMonthTemp=22.197499999999998)
Row(year=u'1994', month=u'07', station=u'83270', AvgMonthTemp=22.177419354838708)
Row(year=u'1994', month=u'07', station=u'103080', AvgMonthTemp=22.164516129032254)
Row(year=u'1994', month=u'07', station=u'82110', AvgMonthTemp=22.161290322580644)
Row(year=u'1994', month=u'07', station=u'97120', AvgMonthTemp=22.13548387096774)
Row(year=u'2010', month=u'07', station=u'98210', AvgMonthTemp=22.111290322580647)
Row(year=u'1994', month=u'07', station=u'53430', AvgMonthTemp=22.096774193548388)
Row(year=u'1997', month=u'08', station=u'86330', AvgMonthTemp=22.079032258064515)
Row(year=u'2006', month=u'07', station=u'66500', AvgMonthTemp=22.054838709677423)
Row(year=u'1994', month=u'07', station=u'76530', AvgMonthTemp=22.033870967741933)
Row(year=u'1997', month=u'08', station=u'98210', AvgMonthTemp=21.983870967741936)
Row(year=u'2014', month=u'07', station=u'98210', AvgMonthTemp=21.962903225806453)
Row(year=u'1994', month=u'07', station=u'62400', AvgMonthTemp=21.951612903225804)
```

```
Row(year=u'1997', month=u'08', station=u'62400', AvgMonthTemp=21.938709677419357)
Row(year=u'1994', month=u'07', station=u'108110', AvgMonthTemp=21.908064516129027)
Row(year=u'1994', month=u'07', station=u'83130', AvgMonthTemp=21.9)
Row(year=u'1997', month=u'08', station=u'98180', AvgMonthTemp=21.88709677419355)
Row(year=u'2006', month=u'07', station=u'98210', AvgMonthTemp=21.872580645161285)
Row(year=u'1997', month=u'08', station=u'98290', AvgMonthTemp=21.864516129032257)
Row(year=u'1991', month=u'08', station=u'78040', AvgMonthTemp=21.85)
Row(year=u'2010', month=u'07', station=u'78140', AvgMonthTemp=21.83064516129032)
Row(year=u'1994', month=u'07', station=u'63340', AvgMonthTemp=21.7758064516129)
Row(year=u'1994', month=u'07', station=u'91130', AvgMonthTemp=21.76290322580645)
Row(year=u'1994', month=u'07', station=u'105370', AvgMonthTemp=21.761290322580646)
Row(year=u'2008', month=u'07', station=u'83420', AvgMonthTemp=21.75)
Row(year=u'1994', month=u'07', station=u'64130', AvgMonthTemp=21.72741935483871)
Row(year=u'1997', month=u'08', station=u'96550', AvgMonthTemp=21.725806451612904)
Row(year=u'1994', month=u'07', station=u'83440', AvgMonthTemp=21.71612903225807)
Row(year=u'2006', month=u'07', station=u'85210', AvgMonthTemp=21.706451612903226)
Row(year=u'1994', month=u'07', station=u'74420', AvgMonthTemp=21.690322580645162)
Row(year=u'1997', month=u'08', station=u'52230', AvgMonthTemp=21.685483870967747)
Row(year=u'2003', month=u'07', station=u'98180', AvgMonthTemp=21.68548387096774)
Row(year=u'1997', month=u'08', station=u'63340', AvgMonthTemp=21.682258064516123)
Row(year=u'1994', month=u'07', station=u'98210', AvgMonthTemp=21.66612903225807)
Row(year=u'1997', month=u'08', station=u'87400', AvgMonthTemp=21.666129032258063)
Row(year=u'1997', month=u'08', station=u'54300', AvgMonthTemp=21.661290322580648)
Row(year=u'1994', month=u'07', station=u'98290', AvgMonthTemp=21.661290322580644)
Row(year=u'2002', month=u'08', station=u'97490', AvgMonthTemp=21.65483870967742)
Row(year=u'2006', month=u'07', station=u'98180', AvgMonthTemp=21.651612903225804)
Row(year=u'2002', month=u'08', station=u'66500', AvgMonthTemp=21.643548387096775)
Row(year=u'1997', month=u'08', station=u'85210', AvgMonthTemp=21.604838709677423)
Row(year=u'2003', month=u'07', station=u'98210', AvgMonthTemp=21.603225806451615)
Row(year=u'1994', month=u'07', station=u'107130', AvgMonthTemp=21.595161290322583)
Row(year=u'1975', month=u'08', station=u'53560', AvgMonthTemp=21.58709677419355)
Row(year=u'1997', month=u'08', station=u'73090', AvgMonthTemp=21.56774193548387)
Row(year=u'2006', month=u'07', station=u'53430', AvgMonthTemp=21.564516129032267)
Row(year=u'2003', month=u'07', station=u'97490', AvgMonthTemp=21.548387096774192)
Row(year=u'1968', month=u'07', station=u'95230', AvgMonthTemp=21.53333333333333)
Row(year=u'1994', month=u'07', station=u'74530', AvgMonthTemp=21.522580645161284)
Row(year=u'2002', month=u'08', station=u'87400', AvgMonthTemp=21.503225806451614)
Row(year=u'1997', month=u'08', station=u'66500', AvgMonthTemp=21.496774193548386)
Row(year=u'1994', month=u'07', station=u'54330', AvgMonthTemp=21.469354838709673)
Row(year=u'1997', month=u'08', station=u'96000', AvgMonthTemp=21.448387096774198)
Row(year=u'1997', month=u'08', station=u'87150', AvgMonthTemp=21.440322580645155)
Row(year=u'1994', month=u'07', station=u'106040', AvgMonthTemp=21.427419354838708)
Row(year=u'1972', month=u'07', station=u'98210', AvgMonthTemp=21.42096774193549)
Row(year=u'1973', month=u'07', station=u'98210', AvgMonthTemp=21.419354838709683)
Row(year=u'1994', month=u'07', station=u'117440', AvgMonthTemp=21.41774193548387)
Row(year=u'2014', month=u'07', station=u'108110', AvgMonthTemp=21.396774193548385)
Row(year=u'1994', month=u'07', station=u'62560', AvgMonthTemp=21.38709677419355)
Row(year=u'1973', month=u'07', station=u'97150', AvgMonthTemp=21.387096774193544)
```

4.Provide a list of stations with their associated maximum measured temperatures and maximum measured daily precipitation. Show only those stations where the maximum temperature is between 25 and 30 degrees and maximum daily precipitation is between 100mm and 200mm.(1950-2014)

station, maxTemp, maxDailyPrecipitation ORDER BY station DESC

(hint: the correct result for this quetion should be empty)

```python
from pyspark import SparkContext
from pyspark.sql import functions as F
from pyspark.sql import SQLContext, Row

sc = SparkContext(appName="BDA2_Q4")
sqlContext = SQLContext(sc)

###Read the data
dataTemp = sc.textFile("/user/x_saeju/data/temperature-readings.csv").cache()
dataTemp = dataTemp.map(lambda line:line.split(";"))
dataTemp = dataTemp.filter(lambda x:(int(x[1][0:4])>=1950 and int(x[1][0:4])<=2014))
dataTemp = dataTemp.map(lambda x:Row(station=x[0], date=x[1], year=x[1].split("-")[0], month=x[1].split
dataTemp = sqlContext.createDataFrame(dataTemp)
dataTemp.registerTempTable("Q3_dataTemp")

dataPrec = sc.textFile("/user/x_saeju/data/precipitation-readings.csv").cache()
dataPrec = dataPrec.map(lambda line:line.split(";"))
dataPrec = dataPrec.filter(lambda x:(int(x[1][0:4])>=1950 and int(x[1][0:4])<=2014))
dataPrec = dataPrec.map(lambda x:Row(station=x[0], date=x[1], year=x[1].split("-")[0], month=x[1].split
dataPrec = sqlContext.createDataFrame(dataPrec)
dataPrec.registerTempTable("Q3_dataPrec")

###Find the maximum temperature for each station
maxTemp = dataTemp.groupBy("station").agg(F.max("temp").alias("maxTemp"))

###Find the maximum "DAILY" precipitation for each station
#First, sum the hourly measured precipitiation for daily precipitation
sumPrecDaily = dataPrec.groupBy("station","date").agg(F.sum("prec").alias("precSum"))
#Then take the maximum Prec for each station
maxPrecDaily = sumPrecDaily.groupBy("station").agg(F.max("precSum").alias("maxPrec"))

###Join the two table by station
#the unmatching date/station will automatically dropped
maxTempPrec = maxTemp.join(maxPrecDaily, "station")


###"25 < maxTemp < 30"
resTEMP = maxTempPrec.filter(maxTempPrec.maxTemp>25).filter(maxTempPrec.maxTemp<30)
resTEMP = resTEMP.orderBy(["station"], ascending=[0])
resTEMP.rdd.repartition(1).saveAsTextFile("Q4_resTEMP")

###"100 < maxPrec < 200"
resPREC = maxTempPrec.filter(maxTempPrec.maxPrec>100).filter(maxTempPrec.maxPrec<200)
resPREC = resPREC.orderBy(["station"], ascending=[0])
resPREC.rdd.repartition(1).saveAsTextFile("Q4_resPREC")
```

```
###"25 < maxTemp < 30" AND "100 < maxPrec < 200"
resTEMPPREC = resTEMP.filter(resTEMP.maxPrec>100).filter(resTEMP.maxPrec<200)
resTEMPPREC = resTEMPPREC.orderBy(["station"], ascending=[0])
resTEMPPREC.rdd.repartition(1).saveAsTextFile("Q4_resTEMPPREC")
```

Results of resTEMP

```
Row(station=u'99280', maxTemp=25.5, maxPrec=34.00000000000001)
Row(station=u'87440', maxTemp=26.2, maxPrec=45.1)
Row(station=u'82360', maxTemp=29.9, maxPrec=84.30000000000001)
Row(station=u'78550', maxTemp=29.9, maxPrec=49.199999999999996)
Row(station=u'71190', maxTemp=28.3, maxPrec=80.39999999999999)
Row(station=u'68560', maxTemp=28.5, maxPrec=52.99999999999999)
Row(station=u'66110', maxTemp=26.3, maxPrec=72.60000000000002)
Row(station=u'52240', maxTemp=29.0, maxPrec=79.6)
Row(station=u'188790', maxTemp=26.6, maxPrec=34.199999999999996)
Row(station=u'183750', maxTemp=29.4, maxPrec=38.1)
Row(station=u'182910', maxTemp=29.2, maxPrec=55.30000000000002)
Row(station=u'180770', maxTemp=28.9, maxPrec=48.6)
Row(station=u'179960', maxTemp=27.9, maxPrec=54.10000000000001)
Row(station=u'177930', maxTemp=26.5, maxPrec=35.9)
Row(station=u'166910', maxTemp=27.1, maxPrec=65.49999999999999)
Row(station=u'162790', maxTemp=28.2, maxPrec=56.599999999999994)
Row(station=u'155790', maxTemp=28.0, maxPrec=49.199999999999996)
Row(station=u'148040', maxTemp=29.9, maxPrec=44.70000000000001)
Row(station=u'147560', maxTemp=29.9, maxPrec=41.800000000000004)
Row(station=u'140460', maxTemp=29.9, maxPrec=48.699999999999996)
Row(station=u'124300', maxTemp=27.2, maxPrec=54.19999999999996)
Row(station=u'123460', maxTemp=29.0, maxPrec=40.2)
Row(station=u'123060', maxTemp=28.0, maxPrec=60.099999999999994)
Row(station=u'114410', maxTemp=29.1, maxPrec=54.3)
Row(station=u'103420', maxTemp=29.7, maxPrec=62.70000000000001)
Row(station=u'99270', maxTemp=27.8, maxPrec=0.0)
Row(station=u'87140', maxTemp=29.5, maxPrec=53.70000000000001)
Row(station=u'81050', maxTemp=29.0, maxPrec=54.300000000000004)
Row(station=u'78280', maxTemp=28.7, maxPrec=52.2)
Row(station=u'192840', maxTemp=27.9, maxPrec=37.0)
Row(station=u'191910', maxTemp=27.7, maxPrec=40.1)
Row(station=u'189720', maxTemp=28.1, maxPrec=56.199999999999996)
Row(station=u'188850', maxTemp=27.9, maxPrec=70.0)
Row(station=u'169790', maxTemp=28.4, maxPrec=40.6)
Row(station=u'163900', maxTemp=29.9, maxPrec=44.5)
Row(station=u'158740', maxTemp=29.2, maxPrec=61.1)
Row(station=u'157870', maxTemp=29.9, maxPrec=61.29999999999999)
Row(station=u'144310', maxTemp=28.4, maxPrec=53.5)
Row(station=u'139260', maxTemp=29.4, maxPrec=86.30000000000001)
Row(station=u'139120', maxTemp=27.9, maxPrec=95.4)
Row(station=u'135460', maxTemp=29.2, maxPrec=62.1)
Row(station=u'122260', maxTemp=28.5, maxPrec=47.6)
Row(station=u'178860', maxTemp=27.4, maxPrec=44.900000000000006)
Row(station=u'133500', maxTemp=27.2, maxPrec=50.10000000000001)
Row(station=u'132170', maxTemp=28.0, maxPrec=39.6)
```

Results of resPREC(first 100 lines)

```
Row(station=u'97510', maxTemp=32.9, maxPrec=103.99999999999999)
Row(station=u'75250', maxTemp=34.4, maxPrec=101.8)
Row(station=u'71420', maxTemp=33.1, maxPrec=106.3)
Row(station=u'52350', maxTemp=32.7, maxPrec=101.6)
```

Results(first 100 lines)

```
#NONE
```

5.Calculate the average monthly precipitation for the Ostergotland region for the period 1993-2016.

In order to do this, you will first need to caculate the total monthly precipitation for each station before calculating the monthly average(by averaging over stations) year, month, avgMonthlyPrecipitation ORDER BY year DESC, month D

```python
from pyspark import SparkContext
from pyspark.sql import functions as F
from pyspark.sql import SQLContext, Row

sc = SparkContext(appName="BDA2_Q5")
sqlContext = SQLContext(sc)

###Read data
station = sc.textFile("/user/x_saeju/data/stations-Ostergotland.csv").cache()
station = station.map(lambda line:line.split(";"))
station = station.map(lambda x:Row(station=x[0], name=x[1]))
stations = sqlContext.createDataFrame(station)
stations.registerTempTable("Q5_Stations")

dataPrec = sc.textFile("/user/x_saeju/data/precipitation-readings.csv").cache()
dataPrec = dataPrec.map(lambda line:line.split(";"))
dataPrec = dataPrec.filter(lambda x:(int(x[1][0:4])>=1993 and int(x[1][0:4])<=2016))
dataPrec = dataPrec.map(lambda x:Row(station=x[0], date=x[1], year=x[1].split("-")[0],
                                month=x[1].split("-")[1], day=x[1].split("-")[2],
                                time=x[2], prec=float(x[3]), quality=x[4]))
dataPrec = sqlContext.createDataFrame(dataPrec)
dataPrec.registerTempTable("Q5_dataPrec")


###
Result = stations.join(dataPrec, "station")
Result = Result.groupBy("year", "month", "station").agg(F.sum("prec").alias("precSum"))
Result = Result.groupBy("year", "month").agg(F.avg("precSum").alias("precAvg"))
Result = Result.orderBy(["year", "month"], ascending=[0,0])

Result.rdd.repartition(1).saveAsTextFile("Q5_result")
```

Results(first 100 lines)

```
Row(year=u'2016', month=u'05', precAvg=29.250000000000007)
Row(year=u'2016', month=u'02', precAvg=21.5625)
```

```
Row(year=u'2015', month=u'12', precAvg=28.925)
Row(year=u'2015', month=u'11', precAvg=63.88750000000002)
Row(year=u'2015', month=u'09', precAvg=101.29999999999998)
Row(year=u'2015', month=u'08', precAvg=26.9875)
Row(year=u'2015', month=u'07', precAvg=119.09999999999995)
Row(year=u'2015', month=u'06', precAvg=78.66250000000001)
Row(year=u'2015', month=u'05', precAvg=93.225)
Row(year=u'2015', month=u'04', precAvg=15.3375)
Row(year=u'2015', month=u'03', precAvg=42.61250000000001)
Row(year=u'2014', month=u'07', precAvg=22.9875)
Row(year=u'2014', month=u'06', precAvg=75.1375)
Row(year=u'2014', month=u'05', precAvg=58.00000000000001)
Row(year=u'2014', month=u'03', precAvg=36.56250000000001)
Row(year=u'2014', month=u'02', precAvg=43.71250000000001)
Row(year=u'2013', month=u'12', precAvg=42.26250000000002)
Row(year=u'2013', month=u'11', precAvg=46.37500000000003)
Row(year=u'2013', month=u'09', precAvg=26.18750000000001)
Row(year=u'2013', month=u'08', precAvg=54.075)
Row(year=u'2013', month=u'07', precAvg=54.56249999999999)
Row(year=u'2013', month=u'05', precAvg=47.92500000000001)
Row(year=u'2013', month=u'04', precAvg=38.287500000000016)
Row(year=u'2013', month=u'02', precAvg=25.52500000000002)
Row(year=u'2012', month=u'11', precAvg=68.65)
Row(year=u'2012', month=u'10', precAvg=65.58333333333334)
Row(year=u'2012', month=u'09', precAvg=72.75)
Row(year=u'2012', month=u'07', precAvg=59.06666666666667)
Row(year=u'2012', month=u'04', precAvg=62.78333333333335)
Row(year=u'2012', month=u'01', precAvg=43.55000000000003)
Row(year=u'2011', month=u'11', precAvg=13.466666666666667)
Row(year=u'2011', month=u'10', precAvg=43.75)
Row(year=u'2011', month=u'09', precAvg=52.56666666666667)
Row(year=u'2011', month=u'05', precAvg=37.85)
Row(year=u'2011', month=u'01', precAvg=35.13333333333335)
Row(year=u'2010', month=u'12', precAvg=37.18333333333339)
Row(year=u'2010', month=u'11', precAvg=93.54999999999994)
Row(year=u'2010', month=u'09', precAvg=43.08333333333334)
Row(year=u'2010', month=u'08', precAvg=108.05)
Row(year=u'2010', month=u'06', precAvg=48.650000000000006)
Row(year=u'2010', month=u'05', precAvg=67.16666666666669)
Row(year=u'2010', month=u'03', precAvg=23.88333333333334)
Row(year=u'2010', month=u'02', precAvg=52.75000000000005)
Row(year=u'2010', month=u'01', precAvg=35.98333333333334)
Row(year=u'2009', month=u'11', precAvg=64.21666666666665)
Row(year=u'2009', month=u'10', precAvg=56.83333333333338)
Row(year=u'2009', month=u'08', precAvg=61.566666666666684)
Row(year=u'2009', month=u'07', precAvg=113.16666666666663)
Row(year=u'2009', month=u'05', precAvg=54.16666666666668)
Row(year=u'2009', month=u'02', precAvg=24.78333333333335)
Row(year=u'2009', month=u'01', precAvg=15.883333333333328)
Row(year=u'2008', month=u'12', precAvg=43.48333333333338)
Row(year=u'2008', month=u'09', precAvg=47.36666666666668)
Row(year=u'2008', month=u'08', precAvg=138.51666666666657)
Row(year=u'2008', month=u'06', precAvg=42.933333333333344)
```

```
Row(year=u'2008', month=u'05', precAvg=23.133333333333336)
Row(year=u'2008', month=u'04', precAvg=20.25)
Row(year=u'2007', month=u'12', precAvg=54.700000000000024)
Row(year=u'2007', month=u'11', precAvg=50.68333333333339)
Row(year=u'2007', month=u'10', precAvg=28.116666666666674)
Row(year=u'2007', month=u'08', precAvg=54.166666666666664)
Row(year=u'2007', month=u'05', precAvg=40.51666666666667)
Row(year=u'2007', month=u'02', precAvg=33.066666666666684)
Row(year=u'2006', month=u'12', precAvg=29.733333333333334)
Row(year=u'2006', month=u'11', precAvg=71.71666666666668)
Row(year=u'2006', month=u'09', precAvg=19.26666666666667)
Row(year=u'2006', month=u'08', precAvg=148.0833333333333)
Row(year=u'2006', month=u'06', precAvg=31.133333333333344)
Row(year=u'2006', month=u'03', precAvg=27.86666666666667)
Row(year=u'2005', month=u'12', precAvg=56.633333333333404)
Row(year=u'2005', month=u'10', precAvg=38.050000000000004)
Row(year=u'2005', month=u'09', precAvg=13.950000000000001)
Row(year=u'2005', month=u'07', precAvg=104.34999999999997)
Row(year=u'2005', month=u'06', precAvg=67.96666666666665)
Row(year=u'2005', month=u'04', precAvg=11.65)
Row(year=u'2005', month=u'01', precAvg=18.05)
Row(year=u'2004', month=u'09', precAvg=37.20000000000001)
Row(year=u'2004', month=u'06', precAvg=56.85000000000002)
Row(year=u'2004', month=u'03', precAvg=28.483333333333338)
Row(year=u'2004', month=u'01', precAvg=26.40000000000001)
Row(year=u'2003', month=u'12', precAvg=52.116666666666674)
Row(year=u'2003', month=u'10', precAvg=45.58333333333334)
Row(year=u'2003', month=u'09', precAvg=8.883333333333333)
Row(year=u'2003', month=u'07', precAvg=113.46666666666665)
Row(year=u'2003', month=u'04', precAvg=51.416666666666664)
Row(year=u'2003', month=u'01', precAvg=17.716666666666672)
Row(year=u'2002', month=u'09', precAvg=16.066666666666666)
Row(year=u'2002', month=u'06', precAvg=98.7833333333333)
Row(year=u'2002', month=u'04', precAvg=29.91666666666667)
Row(year=u'2002', month=u'03', precAvg=26.93333333333334)
Row(year=u'2002', month=u'01', precAvg=55.00000000000001)
Row(year=u'2001', month=u'12', precAvg=35.183333333333344)
Row(year=u'2001', month=u'10', precAvg=60.48333333333335)
Row(year=u'2001', month=u'07', precAvg=40.28333333333334)
Row(year=u'2001', month=u'04', precAvg=46.06666666666669)
Row(year=u'2001', month=u'02', precAvg=36.76666666666669)
Row(year=u'2001', month=u'01', precAvg=36.416666666666686)
Row(year=u'2000', month=u'11', precAvg=108.11666666666662)
Row(year=u'2000', month=u'10', precAvg=110.29999999999997)
Row(year=u'2000', month=u'07', precAvg=135.8666666666666)
```

6.Compare the average monthly temeprature(find the difference) in the period 1950-2014 for all stations in Ostergotland with long-term monthly averages in the period of 1950-1980. Then make a plot of your results.

The first step is to find the monthly averages for each station. In the next step, you can average over all stations to acquire the average temperature for a specific year and month. This RDD/Data

Frame can be used to compute the ong-term average by averaging over all the years in the interval.

year, month difference ORDER BY year DESC, month DESC

```python
from pyspark import SparkContext
from pyspark.sql import functions as F
from pyspark.sql import SQLContext, Row

sc = SparkContext(appName="BDA2_Q6")
sqlContext = SQLContext(sc)

station = sc.textFile("/user/x_saeju/data/stations-Ostergotland.csv").cache()
station = station.map(lambda line:line.split(";"))
station = station.map(lambda x:Row(station=x[0], name=x[1]))
stations = sqlContext.createDataFrame(station)
stations.registerTempTable("Q6_Stations")

dataTEMP = sc.textFile("/user/x_saeju/data/temperature-readings.csv").cache()
dataTEMP = dataTEMP.map(lambda line: line.split(";"))
dataTEMP = dataTEMP.filter(lambda x:(int(x[1][0:4])>=1950 and int(x[1][0:4])<=2014))
dataTEMP = dataTEMP.map(lambda x:Row(station=x[0], date=x[1], year=x[1].split("-")[0],
                                    month=x[1].split("-")[1], day=x[1].split("-")[2], time=x[2], temp=
dataTEMP = sqlContext.createDataFrame(dataTEMP)
dataTEMP.registerTempTable("Q6_dataTemp")

###Finding the monthly averages for each station.

#####Finding the date averages of temperature,
q6data = stations.join(dataTEMP, "station")
minTemp = q6data.groupBy("year", "month", "day", "station").agg(F.min("temp").alias("minT"))
maxTemp = q6data.groupBy("year", "month", "day", "station").agg(F.max("temp").alias("maxT"))
AvgTemp = minTemp.join(maxTemp,["year", "month", "day", "station"])
AvgTemp = AvgTemp.withColumn("AvgdateTemp",(AvgTemp.minT+AvgTemp.maxT)/2)


#####Now find the MONTHLY average temperature(find the monthly averages for each station)
AvgTemp = AvgTemp.select("year", "month", "station", "AvgdateTemp")
AvgTemp2 = AvgTemp.groupBy("year", "month", "station").agg(F.avg("AvgdateTemp").alias("AvgMonTemp")) #s
#AvgTemp2:year,month,station,AvgMonTemp / average temperature of each station for all given year and mo
AvgTemp3 = AvgTemp2.groupBy("year","month").agg(F.avg("AvgMonTemp").alias("MonthlyAvg")) #monthly avera
AvgTemp3 = AvgTemp3.orderBy(["MonthlyAvg"], ascending=[0])

####Now average over all stations to acquire the average temperature for a specific year and month
AvgTemp4 = AvgTemp2.filter(AvgTemp2.year <= 1980)
AvgTemp4 = AvgTemp4.groupBy("month").agg(F.avg("AvgMonTemp").alias("MonthlyAvg2"))
#AvgTemp3 and AvgTemp4.. what is the difference other than that the year difference

diff = AvgTemp3.join(AvgTemp4, "month")
diff = diff.withColumn("diff", diff.MonthlyAvg-diff.MonthlyAvg2)
#diff.rdd.repartition(1).saveAsTextFile("diffwithoutSelect")
diff = diff.select("year","month","diff").orderBy(["year","month"],ascending=[0,0])
diff.rdd.repartition(1).saveAsTextFile("Q6_difference")

####Save it into csv file
diffout = diff.map(lambda x: "%s;%s;%s" % (x[0],x[1],x[2]))
```

```
diffout.repartition(1).saveAsTextFile("Q6_difference_out")
```

Results(first 100 lines)

```
Row(year=u'2014', month=u'12', diff=1.0755653939963379)
Row(year=u'2014', month=u'11', diff=2.102776159654802)
Row(year=u'2014', month=u'10', diff=1.5102783305404266)
Row(year=u'2014', month=u'09', diff=0.05177129602211217)
Row(year=u'2014', month=u'08', diff=-0.6929960424792991)
Row(year=u'2014', month=u'07', diff=2.0408908885004315)
Row(year=u'2014', month=u'06', diff=-2.082721861471864)
Row(year=u'2014', month=u'05', diff=0.13969312945119228)
Row(year=u'2014', month=u'04', diff=2.016597626752966)
Row(year=u'2014', month=u'03', diff=4.364954964390448)
Row(year=u'2014', month=u'02', diff=5.3462126594669686)
Row(year=u'2014', month=u'01', diff=0.9043703473945395)
Row(year=u'2013', month=u'12', diff=4.131349036876754)
Row(year=u'2013', month=u'11', diff=0.9734882808669227)
Row(year=u'2013', month=u'10', diff=0.7406302366987845)
Row(year=u'2013', month=u'09', diff=-1.01434440645723)
Row(year=u'2013', month=u'08', diff=-0.36499017737665795)
Row(year=u'2013', month=u'07', diff=-0.044739610033293076)
Row(year=u'2013', month=u'06', diff=-0.8195400432900488)
Row(year=u'2013', month=u'05', diff=1.4464233347297828)
Row(year=u'2013', month=u'04', diff=-0.662917524762185)
Row(year=u'2013', month=u'03', diff=-3.7068338919145374)
Row(year=u'2013', month=u'02', diff=0.5727548672591762)
Row(year=u'2013', month=u'01', diff=-0.7240314121362519)
Row(year=u'2012', month=u'12', diff=-3.3256076265315215)
Row(year=u'2012', month=u'11', diff=1.3250034323820734)
Row(year=u'2012', month=u'10', diff=-1.4382985195544506)
Row(year=u'2012', month=u'09', diff=-0.5326544742860158)
Row(year=u'2012', month=u'08', diff=-0.8227614383737265)
Row(year=u'2012', month=u'07', diff=-0.7142410763089551)
Row(year=u'2012', month=u'06', diff=-3.2761309523809548)
Row(year=u'2012', month=u'05', diff=0.5392385839966458)
Row(year=u'2012', month=u'04', diff=-0.45291752476218505)
Row(year=u'2012', month=u'03', diff=4.371025345622121)
Row(year=u'2012', month=u'02', diff=-0.17142336829836902)
Row(year=u'2012', month=u'01', diff=1.5648248928490855)
Row(year=u'2011', month=u'12', diff=2.98782473353478)
Row(year=u'2011', month=u'11', diff=2.730609492988134)
Row(year=u'2011', month=u'10', diff=0.12112877042312409)
Row(year=u'2011', month=u'09', diff=1.181684060025784)
Row(year=u'2011', month=u'08', diff=-0.48727756740598593)
Row(year=u'2011', month=u'07', diff=0.5128850233977928)
Row(year=u'2011', month=u'06', diff=-0.2697381784881827)
Row(year=u'2011', month=u'05', diff=0.033666736489314175)
Row(year=u'2011', month=u'04', diff=3.845415808571148)
Row(year=u'2011', month=u'03', diff=0.646831797235023)
Row(year=u'2011', month=u'02', diff=-1.3356987161944067)
Row(year=u'2011', month=u'01', diff=0.2959832506203455)
Row(year=u'2010', month=u'12', diff=-7.13220586700073)
```

```
Row(year=u'2010', month=u'11', diff=-3.0077238403451987)
Row(year=u'2010', month=u'10', diff=-2.1915545140343538)
Row(year=u'2010', month=u'09', diff=-1.2020797140789021)
Row(year=u'2010', month=u'08', diff=-0.29028343250862676)
Row(year=u'2010', month=u'07', diff=1.9040629315111879)
Row(year=u'2010', month=u'06', diff=-1.572696608946611)
Row(year=u'2010', month=u'05', diff=-0.7904412791509596)
Row(year=u'2010', month=u'04', diff=0.7013754045307437)
Row(year=u'2010', month=u'03', diff=-0.5031349400301015)
Row(year=u'2010', month=u'02', diff=-2.2010005439962335)
Row(year=u'2010', month=u'01', diff=-5.389134051432441)
Row(year=u'2009', month=u'12', diff=-1.9239947233057153)
Row(year=u'2009', month=u'11', diff=2.2104579778366196)
Row(year=u'2009', month=u'10', diff=-2.934736332216171)
Row(year=u'2009', month=u'09', diff=0.5111031236148076)
Row(year=u'2009', month=u'08', diff=-0.14812800728868503)
Row(year=u'2009', month=u'07', diff=-0.3170270000626161)
Row(year=u'2009', month=u'06', diff=-2.6354491341991366)
Row(year=u'2009', month=u'05', diff=0.06519709382612149)
Row(year=u'2009', month=u'04', diff=2.706476414631754)
Row(year=u'2009', month=u'03', diff=0.6465385421030582)
Row(year=u'2009', month=u'02', diff=0.45781980232411135)
Row(year=u'2009', month=u'01', diff=1.2875521655763578)
Row(year=u'2008', month=u'12', diff=1.3911079159904727)
Row(year=u'2008', month=u'11', diff=-0.06514808276944084)
Row(year=u'2008', month=u'10', diff=-0.031950408462503965)
Row(year=u'2008', month=u'09', diff=-1.486064201813388)
Row(year=u'2008', month=u'08', diff=-1.3306793269367763)
Row(year=u'2008', month=u'07', diff=0.1322398621074754)
Row(year=u'2008', month=u'06', diff=-1.2149945887445934)
Row(year=u'2008', month=u'05', diff=0.9368864334590103)
Row(year=u'2008', month=u'04', diff=1.7190521722075118)
Row(year=u'2008', month=u'03', diff=1.1021103896103894)
Row(year=u'2008', month=u'02', diff=5.899423026685957)
Row(year=u'2008', month=u'01', diff=4.810426065869613)
Row(year=u'2007', month=u'12', diff=2.8110492649640797)
Row(year=u'2007', month=u'11', diff=-1.0201480827694411)
Row(year=u'2007', month=u'10', diff=-1.302339704650188)
Row(year=u'2007', month=u'09', diff=-0.9888564096055976)
Row(year=u'2007', month=u'08', diff=-0.2167497081684484)
Row(year=u'2007', month=u'07', diff=-1.384035797716578)
Row(year=u'2007', month=u'06', diff=-0.3019642857142877)
Row(year=u'2007', month=u'05', diff=0.2971124842898991)
Row(year=u'2007', month=u'04', diff=2.4514764146317534)
Row(year=u'2007', month=u'03', diff=4.372491621281944)
Row(year=u'2007', month=u'02', diff=0.7524626594669686)
Row(year=u'2007', month=u'01', diff=3.6396049515001114)
Row(year=u'2006', month=u'12', diff=5.547412901327717)
Row(year=u'2006', month=u'11', diff=2.0162155535941952)
Row(year=u'2006', month=u'10', diff=2.0422431399245937)
Row(year=u'2006', month=u'09', diff=2.6175288717796867)
```

Plot

```r
result <- read.csv("Q6_difference_out.csv", sep=";", header=FALSE)
result[,4] <- as.numeric(paste(result[,1], ".", result[,2], sep=""))
plot(result[,4], result[,3], xlab="year.month", ylab="diff", cex=0.5,
     type="l", col="grey")
```