Project 2 (FINAL) : BigData Analysis Project - Weather

Due: April 24 2017 @ 11:55pm

This project will have you perform Data Analysis and processing using MapReduce / Apache Spark. The project will use the weather dataset from https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/. This project will use only 16 years of data (2000 - 2016) for all the stations starting with US and elements TMAX, TMIN.

Creation of datasets

For the Global Historical Climatology Network (GHCN) weather data:

Bash

```
mkdir data  # if it doesn't exist
cd data
for i in `seq 2000 2016`; do
    wget https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/${i}.csv.gz
    gzip -cd ${i}.csv.gz  | grep -e TMIN -e TMAX | grep ^US > ${i}.csv
done
```

For the GHCN station metadata (latitude and longitude):

Bash

```
cd data
wget ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt
awk 'OFS="," {print $1, $2, $3}' ghcnd-stations.txt > ghcnd-stations.csv
```

The Google Maps geocoding API is used to get the actual city and state name for each weather station; see the function `getcity()` in `weatherstats.py` for details.

Weather measurement metadata

The following information serves as a definition of each field in one line of data covering one station-day. Each field described below is separated by a comma (,) and follows the order presented in this document.

```
ID = 11 character station identification code
YEAR/MONTH/DAY = 8 character date in YYYYMMDD format (e.g. 19860529 = May 29, 1986)
ELEMENT = 4 character indicator of element type
DATA VALUE = 5 character data value for ELEMENT
M-FLAG = 1 character Measurement Flag
Q-FLAG = 1 character Quality Flag
S-FLAG = 1 character Source Flag
OBS-TIME = 4-character time of observation in hour-minute format (i.e. 0700 =7:00 am)
```

See section III of the GHCN-Daily ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt file for an explanation of ELEMENT codes and their units as well as the M-FLAG, Q-FLAGS and S-FLAGS.

The OBS-TIME field is populated with the observation times contained in NOAA/NCDC's Multinetwork Metadata System (MMS).

Requirements

Build a Hadoop map/reduce or Apache Spark analysis that yields the following:

- Average TMIN, TMAX for each year excluding abnormalities or missing data
- Maximum TMAX, Minimum TMIN for each year excluding abnormalities or missing data
- 5 hottest, 5 coldest weather stations for each year excluding abnormalities or missing data
- Hottest and coldest day and corresponding weather stations in the entire dataset

Usage

On OSC's Owens cluster cluster:

Bash

```
# request a cluster node
qsub -I -l nodes=1:ppn=12 -l walltime=04:00:00
module load spark

# change into the directory where this repo was cloned
cd /dir/where/you/cloned/this/repo

pyspark --executor-memory 18G --driver-memory 18G
```

Then at the PySpark prompt:

Python

```
>>> from weatherstats import run, run_whole_dataset

>>> run()

2000
====

Avg min temp = 4.4 deg C
Avg max temp = 17.6 deg C

Hottest station #1: USC00416892 (Pecos, TX 79772) - 37.4 deg C
Hottest station #2: USC00411013 (Brackettville, TX 78832) - 36.9 deg C
Hottest station #3: USC00045502 (Mecca, CA 92254) - 36.8 deg C
Hottest station #4: USC00024761 (Lake Havasu City, AZ 86403) - 35.0 deg C
Hottest station #5: USC00021026 (Buckeye, AZ 85326) - 34.0 deg C

[...]

>>> run_whole_dataset()

Entire dataset (2000-2016)
==========================

  * Loading all datasets into a single DataFrame...
  * Computing coldest station for entire dataset...

Coldest station was USC00501684 (Chicken, AK 99732) on 07 Feb 2008: -57.8 deg C

  * Computing hottest station for entire dataset...

Hottest station was USR0000HKAU (Kula, HI 96790) on 13 Feb 2015: 55.6 deg C
```

```
#!/usr/bin/env python
from __future__ import print_function
from pyspark import SparkContext
from pyspark.sql import SQLContext, Row

sc = SparkContext.getOrCreate() #appName='weatherStats')
sqlc = SQLContext(sc)

include_years = range(2000, 2017)  # all data from 2000-2016


def mkdf(filename):
    """
    Read filename (or glob; things like '20??.csv' will work) and return a
    handle to a PySpark DataFrame
    """
    raw = sc.textFile(filename)
    data = raw.map(lambda x: x.split(','))
    table = data.map(lambda r: Row(sta=r[0], date=r[1], meas=r[2],
                                   degc=int(r[3]), m=r[4], q=r[5], s=r[6],
                                   time=r[7]))
    df = sqlc.createDataFrame(table)
    return df.filter(df.q=='')  # prune measurements w/ quality problems


def mkstations(filename):
    """
    Read in comma-separated data with station identifiers and return a PySpark
    DataFrame

    Source: https://mesonet.agron.iastate.edu/sites/networks.php?network=_ALL_&forma
t=csv&nohtml=on

    Headers: stid,station_name,lat,lon,elev,begints,iem_network
    """
    raw = sc.textFile(filename)
    data = raw.map(lambda x: x.split(','))
    table = data.map(lambda r: Row(stid=r[0], lat=r[1], lon=r[2]))
    return sqlc.createDataFrame(table)


def getcity(stations, sta, raw_json=False):
    """
    Given 'stations' DataFrame (generated by mkstations), look up city and
    state from Google Maps API for given station ID 'sta'.
    """
    import json
    from urllib import urlopen
    baseurl = 'https://maps.googleapis.com/maps/api/geocode/json?sensor=false'

    s = stations.filter(stations.stid==sta).first()
    if not s:
        # raise RuntimeError("Station '%s' not found" % sta)
        return '<not found>'

    response = urlopen('%s&latlng=%s,%s' % (baseurl, s.lat, s.lon))
    raw = response.read()
    json = json.loads(raw)

    # FIXME: check status in returned JSON for error(s)
    if raw_json:
        return json
    else:
        # this is sketchy, but it basically (hopefully) discards everything but
        # city, state (ZIP), and country
        a = json['results'][0]['formatted_address']
        return ', '.join(a.split(', ')[-3:-1])


def run():
    """
    Run analyses on individual years in sequence
    """
```



```
    import pyspark.sql.functions as sqlf

    stations = mkstations('data/ghcnd-stations.csv')

    for col in ['begints', 'elev', 'iem_network']: # station_name, stid
        stations = stations.drop(col)

    for year in include_years:
        df = mkdf('data/%s.csv' % str(year))

        print("\n%s\n====\n" % year)

        # Average minimum temperature
        r = df.filter(df.meas=='TMIN').groupBy().avg('degc').first()
        print('Avg min temp = %0.1f deg C' % (r['avg(degc)'] / 10.0))

        # Average maximum temperature
        r = df.filter(df.meas=='TMAX').groupBy().avg('degc').first()
        print('Avg max temp = %0.1f deg C' % (r['avg(degc)'] / 10.0))

        # Five hottest stations (on average)
        # join with 'stations' table (adds lat, lon, station_name, stid)
        fivehot = df.filter(df.meas=='TMAX') \
                    .groupBy(df.sta) \
                    .agg(sqlf.avg('degc')) \
                    .join(stations, df.sta==stations.stid) \
                    .sort(sqlf.desc('avg(degc)')) \
                    .limit(5).collect()

        print()
        i = 1
        for s in fivehot:
            t = float(s['avg(degc)']) / 10.0
            print('Hottest station #%s: %s (%s) - %0.1f deg C'
                  % (i, s.sta, getcity(stations,s.sta), t))
            i = i + 1

        # Five coldest stations (on average)
        fivecold = df.filter(df.meas=='TMIN') \
                     .groupBy(df.sta) \
                     .agg(sqlf.avg('degc')) \
                     .join(stations, df.sta==stations.stid) \
                     .sort(sqlf.asc('avg(degc)')) \
                     .limit(5).collect()

        print()
        i = 1
        for s in fivecold:
            t = float(s['avg(degc)']) / 10.0
            print('Coldest station #%s: %s (%s) - %0.1f deg C'
                  % (i, s.sta, getcity(stations,s.sta), t))
            i = i + 1


def run_whole_dataset():
    """
    Run analyses over the entire dataset
    """
    import pyspark.sql.functions as sqlf
    from datetime import datetime as dt

    stations = mkstations('data/ghcnd-stations.csv')

    # Hottest and coldest day and corresponding weather stations in the
    # entire dataset
    print("\nEntire dataset (2000-2016)\n=========================\n")
    print('  * Loading all datasets into a single DataFrame...')
    df = mkdf('data/20??.csv')

    print('  * Computing coldest station for entire dataset...\n')
    coldest = df.filter(df.meas=='TMIN').groupBy('sta', 'date').min('degc') \
                .sort(sqlf.asc('min(degc)')).first()
```

```
    date = dt.strptime(coldest.date, '%Y%m%d').strftime('%d %b %Y')
    city = getcity(stations, coldest.sta)

    print('Coldest station was %s (%s) on %s: %0.1f deg C'
          % (coldest.sta, city, date, float(coldest['min(degc)']) / 10.0))

    # and now the hottest
    print('\n  * Computing hottest station for entire dataset...\n')
    hottest = df.filter(df.meas=='TMAX').groupBy('sta', 'date').max('degc') \
                .sort(sqlf.desc('max(degc)')).first()

    date = dt.strptime(hottest.date, '%Y%m%d').strftime('%d %b %Y')
    city = getcity(stations, hottest.sta)

    print('Hottest station was %s (%s) on %s: %0.1f deg C'
          % (hottest.sta, city, date, float(hottest['max(degc)']) / 10.0))


if __name__ == '__main__':
    run()
```

```
w687009@owens [~/devel/cs5165/p02]$ pyspark --executor-memory 18G --driver-memory 18
G
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.0.0
      /_/

Using Python version 2.7.5 (default, Aug  2 2016 04:20:16)
SparkSession available as 'spark'.
Python 2.7.5 (default, Aug  2 2016, 04:20:16)
Type "copyright", "credits" or "license" for more information.

IPython 3.2.1 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: %logstart -o
Activating auto-logging. Current session state plus future input saved.
Filename       : ipython_log.py
Mode           : rotate
Output logging : True
Raw input log  : False
Timestamping   : False
State          : active

In [2]: from weatherstats import mkdf, mkstations, getcity, run, run_whole_dataset

In [3]: run()

2000
====

Avg min temp = 4.4 deg C
Avg max temp = 17.6 deg C

Hottest station #1: USC00416892 (Pecos, TX 79772) - 37.4 deg C
Hottest station #2: USC00411013 (Brackettville, TX 78832) - 36.9 deg C
Hottest station #3: USC00045502 (Mecca, CA 92254) - 36.8 deg C
Hottest station #4: USC00024761 (Lake Havasu City, AZ 86403) - 35.0 deg C
Hottest station #5: USC00021026 (Buckeye, AZ 85326) - 34.0 deg C

Coldest station #1: USC00508140 (Salcha, AK 99714) - -20.8 deg C
Coldest station #2: USC00505873 (Fairbanks, AK 99712) - -18.7 deg C
Coldest station #3: USR0000WBAR (Iron River, WI 54847) - -18.1 deg C
Coldest station #4: USC00247248 (Rudyard, MT 59540) - -16.0 deg C
Coldest station #5: USW00026508 (North Slope, AK) - -15.7 deg C

2001
====

Avg min temp = 4.8 deg C
Avg max temp = 17.9 deg C

Hottest station #1: USC00045502 (Mecca, CA 92254) - 35.6 deg C
Hottest station #2: USC00022434 (Dateland, AZ 85333) - 35.1 deg C
Hottest station #3: USC00026250 (Parker, AZ 85344) - 34.0 deg C
Hottest station #4: USC00024761 (Lake Havasu City, AZ 86403) - 33.9 deg C
Hottest station #5: USC00042319 (Furnace Creek, CA 92328) - 33.8 deg C

Coldest station #1: USR0000AHAY (Haycock, AK) - -31.7 deg C
Coldest station #2: USW00026508 (North Slope, AK) - -29.9 deg C
Coldest station #3: USC00508409 (Sutton, AK 99674) - -18.7 deg C
Coldest station #4: USW00026440 (Tanacross, AK 99776) - -18.5 deg C
Coldest station #5: USC00509315 (Tok, AK 99780) - -18.4 deg C

2002
====

Avg min temp = 4.7 deg C
```



```
Avg max temp = 17.7 deg C

Hottest station #1: USC00022807 (Eloy, AZ 85131) - 34.6 deg C
Hottest station #2: USC00347254 (Poteau, OK 74953) - 34.4 deg C
Hottest station #3: USC00042319 (Furnace Creek, CA 92328) - 33.9 deg C
Hottest station #4: USC00044259 (Indio, CA 92201) - 33.4 deg C
Hottest station #5: USR0000CBUU (Winterhaven, CA 92283) - 32.8 deg C

Coldest station #1: USR0000NHAM (Hampden, ND 58338) - -19.2 deg C
Coldest station #2: USS0051R01S (Bettles, AK 99726) - -19.0 deg C
Coldest station #3: USC00502873 (Fairbanks, AK 99709) - -17.3 deg C
Coldest station #4: USC00501987 (Central, AK 99730) - -15.1 deg C
Coldest station #5: USC00263101 (Pioche, NV 89043) - -14.8 deg C

2003
====

Avg min temp = 4.9 deg C
Avg max temp = 17.7 deg C

Hottest station #1: USC00344766 (Kenton, OK 73946) - 36.2 deg C
Hottest station #2: USC00026194 (Gila Bend, AZ 85337) - 35.1 deg C
Hottest station #3: USC00045502 (Mecca, CA 92254) - 34.7 deg C
Hottest station #4: USR0000CWIL (Fillmore, CA 93015) - 34.6 deg C
Hottest station #5: USC00412906 (Encinal, TX 78019) - 34.3 deg C

Coldest station #1: USS0045M07S (Valdez, AK 99686) - -20.1 deg C
Coldest station #2: USS0048V01S (Anchorage, AK 99519) - -18.0 deg C
Coldest station #3: USC00433581 (Groton, VT 05046) - -16.6 deg C
Coldest station #4: USC00067373 (Warren, CT 06777) - -16.3 deg C
Coldest station #5: USC00503181 (Fairbanks, AK 99712) - -15.7 deg C

2004
====

Avg min temp = 5.0 deg C
Avg max temp = 17.5 deg C

Hottest station #1: USC00042346 (Delano, CA 93215) - 37.1 deg C
Hottest station #2: USW00053139 (Stovepipe Wells, CA 92328) - 34.8 deg C
Hottest station #3: USR0000CMIO (Exeter, CA 93221) - 34.3 deg C
Hottest station #4: USW00023195 (Yuma, AZ 85365) - 33.7 deg C
Hottest station #5: USC00044259 (Indio, CA 92201) - 33.4 deg C

Coldest station #1: USR0000NLIM (Bowbells, ND 58721) - -21.0 deg C
Coldest station #2: USC00306957 (Saranac Lake, NY 12983) - -19.7 deg C
Coldest station #3: USC00323686 (Granville, ND 58741) - -18.0 deg C
Coldest station #4: USC00505873 (Fairbanks, AK 99712) - -17.6 deg C
Coldest station #5: USC00505889 (Fairbanks, AK 99712) - -17.5 deg C

2005
====

Avg min temp = 5.0 deg C
Avg max temp = 17.8 deg C

Hottest station #1: USR0000MKIL (Custer, MT 59024) - 34.0 deg C
Hottest station #2: USC00290525 (Portales, NM 88130) - 34.0 deg C
Hottest station #3: USC00427606 (Salt Lake City, UT 84150) - 33.5 deg C
Hottest station #4: USC00383906 (Hampton, SC 29924) - 32.9 deg C
Hottest station #5: USC00311515 (Carthage, NC 28327) - 32.6 deg C

Coldest station #1: USC00305925 (Chestertown, NY 12817) - -27.8 deg C
Coldest station #2: USC00509891 (North Pole, AK 99705) - -19.2 deg C
Coldest station #3: USC00215392 (Milaca, MN 56353) - -18.6 deg C
Coldest station #4: USC00501244 (Cantwell, AK 99729) - -18.5 deg C
Coldest station #5: USC00503585 (Healy, AK 99743) - -17.9 deg C

2006
====

Avg min temp = 5.1 deg C
Avg max temp = 18.1 deg C
```

Hottest station #1: USC00021514 (Queen Creek, AZ 85142) - 40.8 deg C
Hottest station #2: USC00419122 (Trent, TX 79561) - 36.2 deg C
Hottest station #3: USC00080611 (Belle Glade, FL 33430) - 35.0 deg C
Hottest station #4: USR0000HMOL (Kilauea, HI 96754) - 34.0 deg C
Hottest station #5: USC00042319 (Furnace Creek, CA 92328) - 33.4 deg C

Coldest station #1: USC00505534 (Fairbanks, AK 99790) - -24.2 deg C
Coldest station #2: USC00508130 (Anchorage, AK 99519) - -17.8 deg C
Coldest station #3: USC00210059 (Aitkin, MN 56431) - -17.2 deg C
Coldest station #4: USS0049T03S (Dalton Hwy, Alaska) - -17.2 deg C
Coldest station #5: USS0045Q05S (Central, AK 99730) - -16.4 deg C

2007
====

Avg min temp = 4.9 deg C
Avg max temp = 17.8 deg C

Hottest station #1: USC00314987 (Lillington, NC 27546) - 38.3 deg C
Hottest station #2: USC00029376 (Willow Beach, AZ 86445) - 35.5 deg C
Hottest station #3: USR0000HMOL (Kilauea, HI 96754) - 34.6 deg C
Hottest station #4: USC00042319 (Furnace Creek, CA 92328) - 34.0 deg C
Hottest station #5: USC00041048 (Brawley, CA 92227) - 33.8 deg C

Coldest station #1: USC00502707 (Eielson AFB, AK 99702) - -26.0 deg C
Coldest station #2: USC00248857 (West Yellowstone, MT 59758) - -23.3 deg C
Coldest station #3: USC00509315 (Tok, AK 99780) - -20.0 deg C
Coldest station #4: USC00502107 (Fairbanks, AK 99709) - -17.1 deg C
Coldest station #5: USC00506157 (Wasilla, AK 99623) - -16.6 deg C

2008
====

Avg min temp = 4.1 deg C
Avg max temp = 17.0 deg C

Hottest station #1: USC00411671 (Gonzales, TX 78629) - 35.1 deg C
Hottest station #2: USC00254113 (Indianola, NE 69034) - 35.0 deg C
Hottest station #3: USW00003125 (Yuma, AZ 85365) - 34.9 deg C
Hottest station #4: USR0000HMOL (Kilauea, HI 96754) - 34.5 deg C
Hottest station #5: USC00044259 (Indio, CA 92201) - 34.4 deg C

Coldest station #1: USS0051R01S (Bettles, AK 99726) - -26.0 deg C
Coldest station #2: USC00509315 (Tok, AK 99780) - -24.4 deg C
Coldest station #3: USC00390223 (Buffalo, SD 57720) - -19.7 deg C
Coldest station #4: USC00391917 (Conde, SD 57434) - -19.4 deg C
Coldest station #5: USC00476838 (Prairie Du Sac, WI 53578) - -19.3 deg C

2009
====

Avg min temp = 4.4 deg C
Avg max temp = 16.9 deg C

Hottest station #1: USC00141408 (Centralia, KS 66415) - 37.8 deg C
Hottest station #2: USR0000HMOL (Kilauea, HI 96754) - 36.2 deg C
Hottest station #3: USC00144530 (Larned, KS 67550) - 36.1 deg C
Hottest station #4: USC00027370 (Sacaton, AZ 85147) - 34.5 deg C
Hottest station #5: USC00046198 (Niland, CA 92257) - 34.3 deg C

Coldest station #1: USC00501492 (Fairbanks, AK 99701) - -33.4 deg C
Coldest station #2: USC00471618 (Hayward, WI 54843) - -30.5 deg C
Coldest station #3: USC00509315 (Tok, AK 99780) - -27.5 deg C
Coldest station #4: USC00502015 (Nenana, AK 99760) - -24.4 deg C
Coldest station #5: USC00212916 (Fosston, MN 56542) - -23.4 deg C

2010
====

Avg min temp = 4.7 deg C
Avg max temp = 17.2 deg C

Hottest station #1: USC00141408 (Centralia, KS 66415) - 35.0 deg C
Hottest station #2: USC00417951 (San Augustine, TX 75972) - 34.9 deg C
Hottest station #3: USC00027370 (Sacaton, AZ 85147) - 33.4 deg C
Hottest station #4: USC00028070 (Somerton, AZ 85350) - 33.3 deg C
Hottest station #5: USC00341544 (Broken Bow, OK 74728) - 33.1 deg C

Coldest station #1: USC00507097 (Delta Junction, AK 99737) - -24.0 deg C
Coldest station #2: USR0000MKIL (Custer, MT 59024) - -22.9 deg C
Coldest station #3: USC00505873 (Fairbanks, AK 99712) - -22.8 deg C
Coldest station #4: USR0000MOSC (Roy, MT 59471) - -22.7 deg C
Coldest station #5: USR0000MBRV (Geyser, MT 59447) - -21.8 deg C

2011
====

Avg min temp = 4.6 deg C
Avg max temp = 17.3 deg C

Hottest station #1: USR0000CTAR (Camp Pendleton North, CA 92055) - 37.8 deg C
Hottest station #2: USC00411416 (Marfa, TX 79843) - 37.0 deg C
Hottest station #3: USC00035514 (Ozone, AR 72854) - 36.1 deg C
Hottest station #4: USC00412350 (De Leon, TX 76444) - 34.4 deg C
Hottest station #5: USR0000HMOL (Kilauea, HI 96754) - 34.4 deg C

Coldest station #1: USC00479012 (Webster, WI 54893) - -34.4 deg C
Coldest station #2: USC00509314 (Tok, AK 99780) - -24.6 deg C
Coldest station #3: USR0000MOSC (Roy, MT 59471) - -23.0 deg C
Coldest station #4: USR0000MBRV (Geyser, MT 59447) - -22.7 deg C
Coldest station #5: USC00505136 (Kuparuk Rd, Alaska) - -22.6 deg C

2012
====

Avg min temp = 5.4 deg C
Avg max temp = 18.4 deg C

Hottest station #1: USW00012946 (Corpus Christi, TX 78415) - 35.9 deg C
Hottest station #2: USC00290915 (Bosque, NM 87006) - 35.4 deg C
Hottest station #3: USC00412906 (Encinal, TX 78019) - 34.8 deg C
Hottest station #4: USC00040924 (Blythe, CA 92225) - 34.6 deg C
Hottest station #5: USC00042319 (Furnace Creek, CA 92328) - 34.5 deg C

Coldest station #1: USC00504210 (Eagle, AK 99738) - -28.3 deg C
Coldest station #2: USC00504971 (Healy, AK 99743) - -27.5 deg C
Coldest station #3: USC00503368 (Fairbanks, AK 99709) - -22.8 deg C
Coldest station #4: USC00508156 (Salcha, AK 99714) - -22.3 deg C
Coldest station #5: USR0000MKIL (Custer, MT 59024) - -22.0 deg C

2013
====

Avg min temp = 4.4 deg C
Avg max temp = 16.8 deg C

Hottest station #1: USC00260125 (Boulder City, NV 89005) - 36.9 deg C
Hottest station #2: USC00042319 (Furnace Creek, CA 92328) - 33.7 deg C
Hottest station #3: USC00042410 (Desert Center, CA 92239) - 33.6 deg C
Hottest station #4: USC00025270 (Maricopa, AZ 85138) - 33.3 deg C
Hottest station #5: USC00021050 (Bullhead City, AZ 86429) - 33.0 deg C

Coldest station #1: USR0000ASLC (Fairbanks, AK 99712) - -35.7 deg C
Coldest station #2: USC00201940 (Daggett, MI 49821) - -24.4 deg C
Coldest station #3: USR0000NECH (Pine Bluffs, NE 82082) - -22.3 deg C
Coldest station #4: USR0000NFX1 (Dix, NE 69133) - -21.8 deg C
Coldest station #5: USR0000MBRV (Geyser, MT 59447) - -21.6 deg C

2014
====

Avg min temp = 4.4 deg C
Avg max temp = 16.9 deg C

Hottest station #1: USC00413618 (Goliad, TX 77963) - 36.2 deg C

```
Hottest station #2: USC00042319 (Furnace Creek, CA 92328) - 34.7 deg C
Hottest station #3: USC00040924 (Blythe, CA 92225) - 34.1 deg C
Hottest station #4: USW00003145 (Yuma, AZ 85365) - 33.8 deg C
Hottest station #5: USC00029656 (Yuma, AZ 85364) - 33.6 deg C

Coldest station #1: USC00238456 (Troy, MO 63379) - -25.5 deg C
Coldest station #2: USC00210746 (Big Falls, MN 56627) - -24.9 deg C
Coldest station #3: USC00210050 (Middle River, MN 56737) - -24.3 deg C
Coldest station #4: USC00320450 (Valley City, ND 58072) - -23.9 deg C
Coldest station #5: USC00208680 (Watersmeet, MI 49969) - -23.3 deg C

2015
====

Avg min temp = 5.4 deg C
Avg max temp = 17.8 deg C

Hottest station #1: USC00092593 (Bainbridge, GA 39817) - 34.6 deg C
Hottest station #2: USC00403379 (Gallatin, TN 37066) - 34.6 deg C
Hottest station #3: USC00414278 (Coleman, TX 76834) - 34.6 deg C
Hottest station #4: USC00040924 (Blythe, CA 92225) - 34.4 deg C
Hottest station #5: USC00042319 (Furnace Creek, CA 92328) - 34.1 deg C

Coldest station #1: USC00158551 (West Liberty, KY 41472) - -32.2 deg C
Coldest station #2: USC00210050 (Middle River, MN 56737) - -31.9 deg C
Coldest station #3: USC00242347 (Denton, MT 59430) - -24.2 deg C
Coldest station #4: USC00508130 (Anchorage, AK 99519) - -21.8 deg C
Coldest station #5: USC00503212 (Galena, AK 99741) - -20.0 deg C

2016
====

Avg min temp = 5.6 deg C
Avg max temp = 18.0 deg C

Hottest station #1: USC00412906 (Encinal, TX 78019) - 40.5 deg C
Hottest station #2: USC00415101 (Colorado City, TX 79512) - 35.6 deg C
Hottest station #3: USC00406340 (Mt Pleasant, TN 38474) - 34.3 deg C
Hottest station #4: USC00025700 (Mesa, AZ 85215) - 34.3 deg C
Hottest station #5: USC00029656 (Yuma, AZ 85364) - 34.1 deg C

Coldest station #1: USC00200234 (Felch, MI 49831) - -26.6 deg C
Coldest station #2: USC00507778 (Dalton Hwy, Alaska) - -18.5 deg C
Coldest station #3: USC00210746 (Big Falls, MN 56627) - -17.1 deg C
Coldest station #4: USC00246685 (Bonner, MT 59823) - -13.8 deg C
Coldest station #5: USC00509869 (Fairbanks, AK 99701) - -13.5 deg C

In [4]: run_whole_dataset()

Entire dataset (2000-2016)
==========================

  * Loading all datasets into a single DataFrame...
  * Computing coldest station for entire dataset...

Coldest station was USC00501684 (Chicken, AK 99732) on 07 Feb 2008: -57.8 deg C

  * Computing hottest station for entire dataset...

Hottest station was USR0000HKAU (Kula, HI 96790) on 13 Feb 2015: 55.6 deg C

In [5]:
```