Project 1: Exploring Weather Trends

Student´s name: Hernán Adasme Herrera

I. Steps taken to extract the data from SQL to be visualized in Excel:

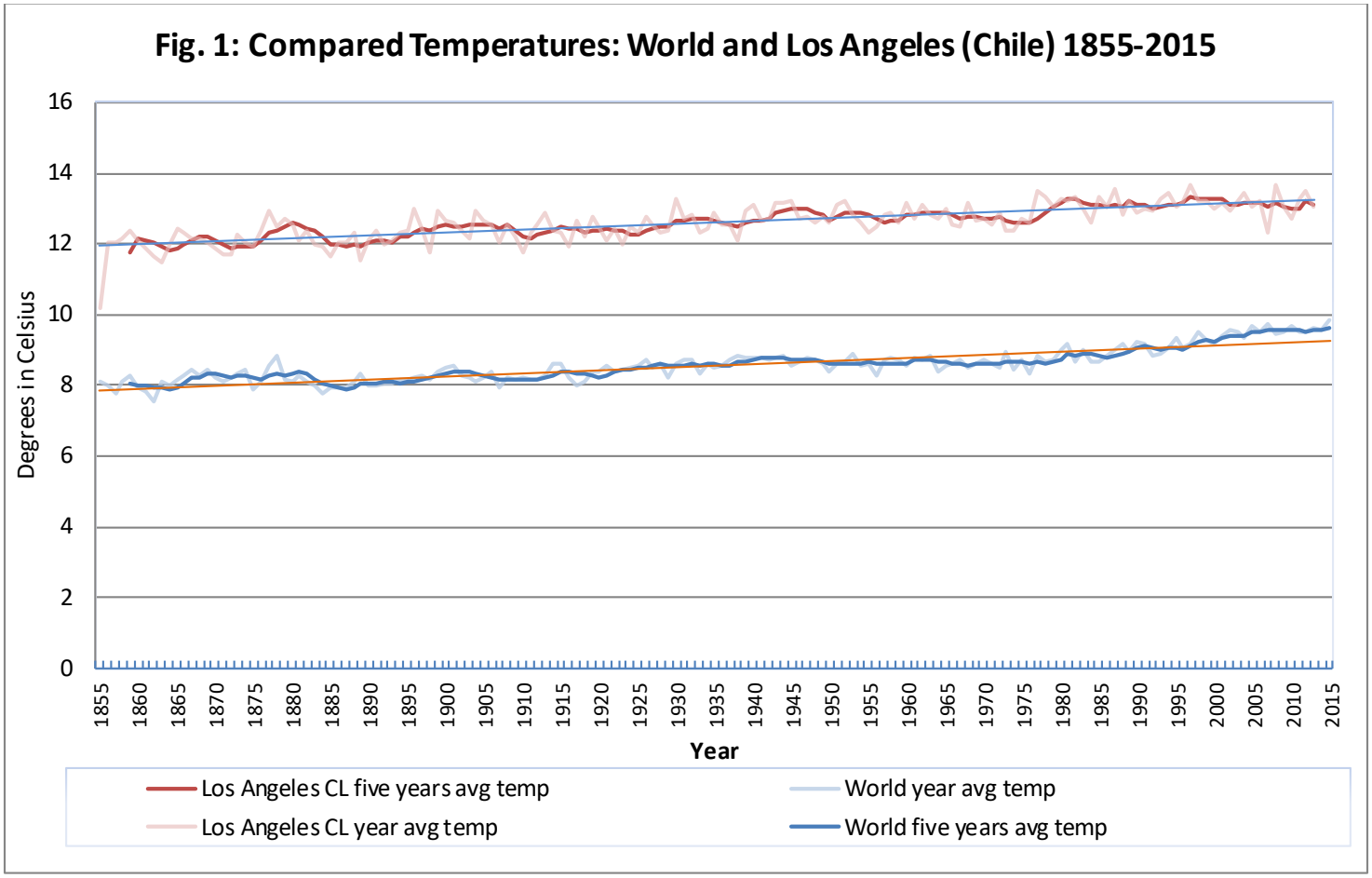
1. Tools: Sql (Udacity Workspace version)
2. SELECT city FROM city_list WHERE country = 'Chile'

   SELECT year, city, country, avg_temp FROM city_data WHERE city = 'Los Angeles' AND country = 'Chile'

   SELECT avg_temp, year FROM global_data
3. Downloaded and saved it as Avg_temp_Los_Angeles_1855_2013.csv
4. Downloaded and saved it as Avg_temp_world_1750_2015.csv
5. Changed format from CSV to **.**xlsx of files; open both files.
6. Create two new columns *world_five_year_mavg_temp* (column C) and *los_angeles_five_year_mavg_temp* (column E)
7. Switching all the commas into dots, in order to avoid the *zero division error.*
8. Calculating the moving averages: Type in C6 =AVERAGE(B2 : B6); moving average from 1750 to 2015. Drag down the formula until the end of the data.
9. Calculating the moving averages: Type in E111 AVERAGE(D107:D111); moving averages from 1850 to 2013. Drag down the formula until the end of the data.
10. Line Chart: Creating a line chart by selecting the data and formatting it.
11. Key considerations:

    - Chilean city Los Angeles has been selected because this city is the closest big city to where I live; Santiago, Chile.
    - Moving averages have been computed based on a 5-yr convention; the five-year interval was selected because it smooths data out, while at the same time it keeps peaks and trends somewhat close to the original data. Compared to the 2-yr and the 10-yr moving averages, the 5-ys moving average provides a good balance between asymmetry and silkiness of the data.
    - Since the data-set from Los Angeles starts in year 1850, this year has been selected as the starting point, even though the Global data-set begins in 1750.
    - The Data set from Los Angeles only reaches up until 2013; however, year 2015 has been set as the ending point, because it corresponds with Global data-set ending year.
    - Smoothed values have been computed from the fifth year on; that is why transparent lines and solid lines do not start at the same point.

II. Observations



1) **General overview**. Figure 1 shows temperature series from Los Angeles, a mid-size city located in Chile´s center-south region, and the globe, between 1855 and 2015. The data is presented on an annual year-to-year basis, and in a 5-yr moving average value; solid lines indicate the 5-yr moving average, and transparent lines represent the annual values. Moving averages have been calculated based on a 5-yr interval, in order to smooth out the data and reduce year-to-year fluctuations; the 5-yr interval makes possible to observe temperature long-term trends. At a first glance, global and Los Angeles 5-yr lines show a clear trend toward global warming, although this upward trend is more prominent on world´s data. On the global data series, this upward trend becomes more noticeable around 1920, a period known as the second industrial revolution. Los Angeles year-to-year and 5-yr series indicate more interval variability than the global series, and also a higher average temperature: around 12 and 14 degrees Celsius throughout the entire period, while global temperatures move around 8 and 10 degrees Celsius during the same time-arc.

2) **Temperature peaks:** some temperature peaks observed on the annual data for both series got shifted on the 5-year moving average, even though shifted peaks are way more observable on Los Angeles data. For instance, annual data from Los Angeles suggest a temperature peak in 1876, while the temperature peak shown in the 5-yr corresponds to 1880. Similarly, the temperature decreasing trend beginning in 1876 is only observable in 1880. In general, since the original data is asymmetric, the smoothed out data persistently shifts the peaks and moves the starting point of upward and downward trends. Shifted peaks might have some serious implications for data interpretation; trends starting at certain point according to moving averages, might´ve indeed started several years before.
3) **Proximity between smoothed data and annual data:** Fig. 1 shows that plateaus, peaks and upwards trends from world´s smoothed data follows much more closely the original year-to-year global data than Los Angeles

data-series. Put it simpler; light blue and dark blue lines are more n 'sync than light and dark red lines. This is probably due to the nature of data and the differences between data-collection; world´s data is already an average from annual global temperatures, while Los Angeles data to a local observation. This explains the great difference in symmetry observable on the data.

4) **Correspondence between the ranges of upward trends:** Even though Los Angeles and the World´s temperatures are located in different ranges of the Celsius degree scale, the upward slope of both throughout the entire period is somewhat similar. The line chart on Figure 1 illustrates that both trends are located within a range two degrees Celsius. This demonstrates an important degree of correspondence between local and global warming trends.