Analysis of historical temperatures in Ottawa.

# Introduction

Are the temperatures rose during the last century in Ottawa? How much did the average temperature risen? Do we have any signs of global warming? Over 100 years of data will be used for the analysis. We will concentrate on average daily temperatures for the Ottawa weather station. We will use appropriate methods for the temperatures analysis like: correlation, linear regression, analysis of variance, classification. We will analyze results, provide conclusions and answer the question if we have any signs of global warming for the city of Ottawa.

# Literature Review

1. <http://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=4333&autofwd=1>

This site describes how the normal temperatures were calculated. How the means and totals for every month were achieved. There is also information about standard deviation of mean daily Celsius temperatures. Calculation of annual standard deviation differs from the other annual element calculations (represents the mean SD calculated from annual means for a given station). At most climatological stations, maximum temperature is the highest temperature recorded in a 24-hour period ending in the morning of the next day. The minimum values are for a period of the same length, beginning in the evening of the previous day. Mean temperature is the average of the two.

1. <http://www.ec.gc.ca/sc-cs/default.asp?lang=En&n=7150CD6C-1>

Based on all data from 1889 the National temperatures were studied. Which years were coldest, warmest? Deviations were analyzed. Later Regional temperatures we discussed (Canada is divided into 11 climate regions). When examined on a regional basis, average annual temperatures for 2015 were among the 10 warmest on record since 1948 for 6 of the 11 climate regions, with 3 regions experiencing their warmest year on record. The 6 regions are: North B.C. Mountains/Yukon (warmest at 2.9°C above average); Pacific Coast (warmest at 2.0°C above average); South B.C. Mountains (warmest at 1.9°C above average); Prairies (third-warmest at 2.0°C above average); Mackenzie District (fifth-warmest at 2.3°C above average); and Northwestern Forest (eighth-warmest at 1.9°C above average). Although none of the 11 climate regions experienced an average annual temperature for 2015 that ranked among the 10 coldest since 1948, Atlantic Canada did have a slightly cooler-than-average year at 0.1°C below average

1. Sites with the average temperature listings :

US: <https://temperature.weatherdb.com/>

CANADA: <https://www.currentresults.com/Weather/Canada/Cities/temperature-annual-average.php>

Compare City Climates: <https://outflux.net/weather/noaa/>

How much has the Global Temperature risen since 1880 (from that year data have been collected)?

<https://www2.ucar.edu/climate/faq/how-much-has-global-temperature-risen-last-100-years>

# Dataset

The dataset was generated from <http://climate.weather.gc.ca/historical_data/search_historic_data_e.html> internet site. It comes with over 20 attributes: Min, Max, Average temperature, several additional flags not needed, as well as many columns describing precipitations. For the purpose of this project the following attributes were extracted: Year, Month, Max Temp, Min Temp, Average Temp, Max Extreme Temp, and Min Extreme Temp. In general it is a good quality data. For the period of my interest there is no missing data.

# Approach

Once this is done, explain each of the steps in detail. What are you planning to do in each step or have already done? For example, in the above case you would create subheadings for each of the steps.

## Step 1: Generate csv report from internet site.

## The dataset was generated from <http://climate.weather.gc.ca/historical_data/search_historic_data_e.html> internet site. There are different stations to choose from. The OTTAWA CDA station was chosen, because it has Monthly data from 1890 to 2006.

## Step 2: Prepare working data subset

* Opened csv file in Microsoft Excel and removed manually not needed columns as well as renamed very long column names. Removed “special” characters.
* Loaded data to IBM Netezza database.
* Studied downloaded data. Noticed good quality. Removed manually missing data from year 1889. Calculated some averages. Attached **SQL\_OTT\_MONTHLY** script.
* Loaded data to RStudio. Attached **Prep\_Ottawa\_temp\_dataset.R**

## Step 3: Analysis of the Average monthly temperatures.

First we calculated average May temperatures for the decades. Calculated Linear Regression for the Average May temperature, based on the Year. It will give us a tool to predict average temperatures in the future.

lm(formula = Avg~Year, data=subset(temp\_hist, Month==5 ))

(Intercept) Year

-0.045150 0.006644

Since the Year coefficient is positive, the temperatures are rising slightly.

Analysis reviled also that the significant temperature rise occurred in years seventies (1970-79). Based on that we Calculated Linear Regression for Years < 1970 and >=1970.

lm(formula = Avg~Year, data=subset(temp\_hist, Month==5 & Year < 1970))

(Intercept) Year

32.93969 -0.01053

Here the Year coefficient is negative, so actually temperatures were going down thru the years 1890-1970.

For the years after 1970:

lm(formula = Avg~Year, data=subset(temp\_hist, Month==5 & Year >= 1970))

(Intercept) Year

-9.69673 0.01166

Here the Year coefficient is positive so actually temperatures were going up. Please note that this short period of time (35 years) was so significant to offset temperatures going down thru years 1890-1970 (80 years). Interesting!

Also several plots Regression Lines were generated as well as Residual and fitted values. The **average** error was calculated using the following formula:

sum (abs (temp\_hist5$Avg - (temp\_hist5$Year \* 0.006644 -0.045150)))/nrow(temp\_hist5)

It was from **1.02** degrees to **1.4** degrees depending of the analysed period.

Correlation have been checked too. There is very weak correlation between the Year and Average temperature:

> cor(temp\_hist5$Year, temp\_hist5$Avg, method = "pearson")

[1] 0.1268144

> cor(temp\_hist5$Year, temp\_hist5$Avg, method = "spearman")

[1] 0.1192893

Non Parametric : Wilcoxon rank-sum test

> Group1 = temp\_hist5$Avg[which (Year>=1970 & Year < 1980)]

> Group2 = temp\_hist5$Avg[which (Year>=1900 & Year < 1910)]

> ranks\_Group1 = rank(cbind(Group1, Group2))[1:10]

> ranks\_Group2 = rank(cbind(Group1, Group2))[11:20]

> n1=length(Group1)

> n2=length(Group2)

> U1 = n1\*n2 + ((n1\*(n1+1)/2) - sum(ranks\_Group1))

> U2 = n1\*n2 + ((n1\*(n1+1)/2) - sum(ranks\_Group2))

> U1

[1] 38

> U2

[1] 62

> U= min(U1,U2)

> alpha = 0.05

> critical\_value = qwilcox(alpha/2,n1,n2) - 1

> critical\_value

[1] 23

> U

[1] 38

> # since U > critical\_values - DO NOT REJECT THE NULL HYPOTHESIS - There is no significant difference in temperatures