Assignment 01 - Exploratory Data Analysis

2022-02-03

Due Date

This assignment is due at the end of the day Febraury 2, 2022

Learning Goals

- Download, read, and get familiar with an external dataset.
- Step through the EDA "checklist" presented in class
- Practice making exploratory plots

Assignment Description

We will work with weather data from the Canadian arctic. There is scientific evidence that the climate is warming and that these effects are particularly evident in high latitudes. The primary question you will answer with these data is whether *daily* temperature in Fort Ross, Nunavut has significantly changed from 1938 to 2021.

The data were downloaded from here.

Your assignment should be completed in R markdown.

Steps

Given the formulated question from the assignment description, you will now conduct EDA Checklist items 2-4.

1. First, read in the original 1938 and 2021 data for the Fort Ross, Nunavut weather station. For each of the two datasets, check for import issues (dimensions, headers, footers, variable names and variable types). Check for any data issues (import issues, missing values, data errors) particularly in the key variable we are analyzing. Make sure you write up a summary of all of your findings.
2. Combine the two years of data into one data frame. This will require cleaning the data – keep only necessary data columns and change the names of the key variables so that they are easier to merge.
3. Explore the main question of interest. Calculate summary statistics for each year (e.g. means, medians, variances) and then conduct some basic analyses that enable you to compare both years (e.g. t-test, linear regression). Be sure to show the results and write up explanations of what you observe in these data.
4. Create exploratory plots (boxplots, histograms, time series) and a basic map in leaflet() that shows the locations of the weather stations (make sure to use different colors for the two years).

This homework has been adapted from the case study in Roger Peng's Exploratory Data Analysis with R