Project - Team 10-7

The Outliers: Morgan Pruchniewski, Scott Burstein, Katie Zhou

10/6/2020

Source of the Data:

We downloaded the data from Kaggle, but the data was originally compiled by the Berkeley Earth Data Lab.

When and How It Was Originally Collected:

The Berkeley Earth Data Page collected the data from 16 pre-existing archives. It was updated to Kaggle in 2017.

We tentatively plan to interpolate these time-series datasets with other datasets in order to draw even more interesting conclusions, potentially regarding the correlation between global warming and the prevalence of natural disasters.

The Cases:

(By Time and Location (for `GlobalLandTemperaturesByMajorCity`))

Each observation in `GlobalLandTemperaturesByMajorCity` is a city and its respective land temperature, coordinates, and Country, which will be used to investigate changes in climate over time. There are 239,177 rows in the major city dataset.

(By Time (for `GlobalLandTemperaturesByMajorCity`))

Each observation in `GlobalTemperatures` is a numeric date/time value (described below) and its respective land temperature, max temperature, min temperature, ocean temperature, and relevant uncertainties to said variables, which will be used to investigate changes in climate over time. There are 239,177 rows in the global dataset.

Relevant Variables:

The relevant variables in the `major_city` dataset include dt (date and time of the recorded temperature, a discrete numeric), AverageTemperature (the average temperature for each city, a continuous numeric), city (nominal categorical), country (nominal categorical), latitude (continuous numeric), and longitude (continuous numeric).

The relevant variables in the `global` dataset include dt (date and time of the recorded temperature, a discrete numeric), LandAverageTemperature (the average global land temperature at that time, continuous numeric), LandMaxTemperature (the highest recorded land temperature of that year, a continuous numeric), LandMinTemperature (the lowest recorded land temperature of that year, a continous numeric), and LandAndOceanAverageTemperature (the average of land and ocean temperature averages for that year, a continuous numeric).

```
#read in datasets
major_city <- read_csv("data/GlobalLandTemperaturesByMajorCity.csv")
```

```
## Parsed with column specification:
## cols(
```

```
##   dt = col_date(format = ""),
##   AverageTemperature = col_double(),
##   AverageTemperatureUncertainty = col_double(),
##   City = col_character(),
##   Country = col_character(),
##   Latitude = col_character(),
##   Longitude = col_character()
## )
```

```
global <- read_csv("data/GlobalTemperatures.csv")
```

```
## Parsed with column specification:
## cols(
##   dt = col_date(format = ""),
##   LandAverageTemperature = col_double(),
##   LandAverageTemperatureUncertainty = col_double(),
##   LandMaxTemperature = col_logical(),
##   LandMaxTemperatureUncertainty = col_logical(),
##   LandMinTemperature = col_logical(),
##   LandMinTemperatureUncertainty = col_logical(),
##   LandAndOceanAverageTemperature = col_logical(),
##   LandAndOceanAverageTemperatureUncertainty = col_logical()
## )
```

```
## Warning: 11952 parsing failures.
##  row                            col           expected               actual                            
## 1201 LandMaxTemperature             1/0/T/F/TRUE/FALSE 8.241999999999999   'data/GlobalTemperatures.
## 1201 LandMaxTemperatureUncertainty  1/0/T/F/TRUE/FALSE 1.7380000000000002  'data/GlobalTemperatures.
## 1201 LandMinTemperature             1/0/T/F/TRUE/FALSE -3.2060000000000004 'data/GlobalTemperatures.
## 1201 LandMinTemperatureUncertainty  1/0/T/F/TRUE/FALSE 2.822               'data/GlobalTemperatures.
## 1201 LandAndOceanAverageTemperature 1/0/T/F/TRUE/FALSE 12.832999999999998  'data/GlobalTemperatures.
## .... .............................. .................. ................... ........................
## See problems(...) for more details.
```

```
#summary stats for `major_city`
glimpse(major_city)
```

```
## Rows: 239,177
## Columns: 7
## $ dt                           <date> 1849-01-01, 1849-02-01, 1849-03-01, ...
## $ AverageTemperature           <dbl> 26.704, 27.434, 28.101, 26.140, 25.42...
## $ AverageTemperatureUncertainty <dbl> 1.435, 1.362, 1.612, 1.387, 1.200, 1....
## $ City                         <chr> "Abidjan", "Abidjan", "Abidjan", "Abi...
## $ Country                      <chr> "Côte D'Ivoire", "Côte D'Ivoire", "Cô...
## $ Latitude                     <chr> "5.63N", "5.63N", "5.63N", "5.63N", "...
## $ Longitude                    <chr> "3.23W", "3.23W", "3.23W", "3.23W", "...
```

```
colnames(major_city)
```

```
## [1] "dt"                           "AverageTemperature"
## [3] "AverageTemperatureUncertainty" "City"
## [5] "Country"                      "Latitude"
## [7] "Longitude"
```

```
#summary stats for `global`
glimpse(global)
```

```
## Rows: 3,192
## Columns: 9
## $ dt                                       <date> 1750-01-01, 1750-02-01, ...
## $ LandAverageTemperature                   <dbl> 3.034, 3.083, 5.626, 8.49...
## $ LandAverageTemperatureUncertainty        <dbl> 3.574, 3.702, 3.076, 2.45...
## $ LandMaxTemperature                       <lgl> NA, NA, NA, NA, NA, NA, N...
## $ LandMaxTemperatureUncertainty            <lgl> NA, NA, NA, NA, NA, NA, N...
## $ LandMinTemperature                       <lgl> NA, NA, NA, NA, NA, NA, N...
## $ LandMinTemperatureUncertainty            <lgl> NA, NA, NA, NA, NA, NA, N...
## $ LandAndOceanAverageTemperature           <lgl> NA, NA, NA, NA, NA, NA, N...
## $ LandAndOceanAverageTemperatureUncertainty <lgl> NA, NA, NA, NA, NA, NA, N...
```

```
colnames(global)
```

```
## [1] "dt"
## [2] "LandAverageTemperature"
## [3] "LandAverageTemperatureUncertainty"
## [4] "LandMaxTemperature"
## [5] "LandMaxTemperatureUncertainty"
## [6] "LandMinTemperature"
## [7] "LandMinTemperatureUncertainty"
## [8] "LandAndOceanAverageTemperature"
## [9] "LandAndOceanAverageTemperatureUncertainty"
```

Research Question:

Is there significant support for rising earth surface temperatures, and if so, how has the rate of change in temperatures fluctuated over time?

H_o = There is no statistical evidence of rising earth surface temperature since 1750. Thus, the rate of temperature change is constant (0).

H_a = There is statistical evidence of rising earth surface temperature since 1750, and the rate of annual temperature change has also changed.

How do the changes in earth surface temperature differ geographically, and are there regions where climate change seems to have a greater effect?

H_o = There is no variation in earth surface temperature change geographically, and all regions experience equal effects of global warming. i.e. latitude and longitude values are not correlated with temperature variability since 1750.

H_a = There is variation in earth surface temperature change geographically, and regions experience differing effects of global warming based on their unique geopositions.