Climate change evidence

Contents

1 Data processing

1.1 NASA/GISS data

Analyze global land-ocean data. No need of pre-processing, data already clean.

```
df_temp_anomaly <- read.csv("data/NASA/global_temperature_anomaly.csv",
                            sep = ",", header = TRUE, skip = 2)

# Have a look at first 5 row of the dataset
kable(head(df_temp_anomaly))
```

Year	No_Smoothing	Lowess.5.
1880	-0.16	-0.09
1881	-0.08	-0.12
1882	-0.10	-0.16
1883	-0.17	-0.20
1884	-0.28	-0.23
1885	-0.33	-0.26

1.2 Berkley Earth data

Have a look at different data per continent.

1.2.1 Analysis of one continent file

Do the processing of one single file, i.e. data processing of one continent to have a better idea of what to do

```
# Define columns data frame
cols_month <- c("year", "month", "monthly_anomaly",
                "monthly_uncertainty", "annual_anomaly",
                "annual_uncertainty", "five_year_anomaly",
                "five_year_uncertainty", "ten_year_anomaly",
                "ten_year_uncertainty", "twenty_year_anomaly", "twenty_year_uncertainty")

path <- "data/Berkley Earth/continent_monthly_avg_temp/europe-TAVG-Trend.txt"

# Read data
df_europe_month <- read_delim(path,
                        delim = " ",
                        col_names = cols_month,
                        # Specify all columns as character otherwise error show up
                        col_types = "cccccccccccc",
                        skip = 70,
                        na = c("", "NA","NaN"))

# Only keep first 4 columns
df_europe_month <- df_europe_month[, 1:4]

# Convert all columns to numeric
df_europe_month <- data.frame(sapply(df_europe_month, as.numeric))

# Get name of zone from the path name
zone_name_path <- unlist(strsplit(path, split = "/"))[4]

# Get name of the zone as a list
zone_name <- head(unlist(strsplit(zone_name_path, split = "-")), -2)

# Collapse name into a single string
zone_name <- paste(zone_name, collapse = '_')

# Add name of the zone as a column to the dataset
df_europe_month$zone <- zone_name

# Create a single date column
df_europe_month$date = as.yearmon(paste(df_europe_month$year, df_europe_month$month), "%Y %m")

# Delete column year and month
df_europe_month <- subset(df_europe_month, select = -c(year,month))

dim(df_europe_month)
```

```
## [1] 3252    4
```

```
str(df_europe_month)
```

```
## 'data.frame':    3252 obs. of  4 variables:
```

```
##  $ monthly_anomaly    : num  0.119 0.886 1.855 0.566 -1.065 ...
##  $ monthly_uncertainty: num  1.75 2.28 2.59 1.56 1.47 ...
##  $ zone               : chr  "europe" "europe" "europe" "europe" ...
##  $ date               : 'yearmon' num  Jan 1750 Feb 1750 Mar 1750 Apr 1750 ...
```

```
kable(head(df_europe_month))
```

monthly_anomaly	monthly_uncertainty	zone	date
0.119	1.749	europe	Jan 1750
0.886	2.276	europe	Feb 1750
1.855	2.588	europe	Mar 1750
0.566	1.558	europe	Apr 1750
-1.065	1.472	europe	May 1750
-0.099	1.525	europe	Jun 1750

```
kable(summary(df_europe_month))
```

monthly_anomaly	monthly_uncertainty	zone	date
Min. :-6.3520	Min. : 0.057	Length:3252	Min. :1750
1st Qu.:-0.8530	1st Qu.: 0.210	Class :character	1st Qu.:1818
Median :-0.0540	Median : 0.523	Mode :character	Median :1885
Mean :-0.1167	Mean : 1.237	NA	Mean :1885
3rd Qu.: 0.6930	3rd Qu.: 1.814	NA	3rd Qu.:1953
Max. : 5.4490	Max. :10.288	NA	Max. :2021
NA's :1	NA's :1	NA	NA

1.2.2 Aggregate all data

Using a for loop, read all the files per single continent contained in the same folder and apply the necessary cleaning

```
# Get a list with all the files inside the folder zone_monthly_average_temp
files <- list.files(path="data/Berkley Earth/continent_monthly_avg_temp",
                    pattern="*.txt",
                    full.names=TRUE,
                    recursive=FALSE)

# Create empty data frame with correct columns to store all the data
df_continent_avg_temp <- data.frame(zone = character(),
                                    year = character(),
                                    anomaly = double(),
                                    lowess = double(),
                                    uncertainty = double(),
                                    stringsAsFactors=FALSE)

# Loop true all the files in the folder and read them
for (file in files){

  # Read and create a temporary dataset
```

```
  df_year_temp <- read_delim(file,
                         delim = " ",
                         col_names = cols_month,
                         # Specify all columns as character otherwise error show up
                         col_types = "cccccccccccc",
                         skip = 70,
                         na = c("", "NA","NaN"))

  # Only keep first 4 columns
  df_year_temp <- df_year_temp[, 1:4]

  # Convert all columns to numeric
  df_year_temp <- data.frame(sapply(df_year_temp, as.numeric))

  # CALCULATE AVERAGE PER YEAR GIVEN MONTHLY DATA

  # Aggregate data per year and calculate mean
  df_year_temp <- aggregate(monthly_anomaly~year, transform(df_year_temp, year = df_year_temp$year), mea

  # Change name of the column
  colnames(df_year_temp)[2] <- "yearly_anomaly"

  # Compute LOWESS smoother which uses locally-weighted polynomial regression
  line_trend <- lowess(as.numeric(df_year_temp$year), df_year_temp$yearly_anomaly, f = 1/5)

  # Add lowess to year temp dataframe as new column
  df_year_temp$lowess <- line_trend$y

  # GET ZONE NAME FOR PATH

  # Unlist path name
  zone_name_path <- unlist(strsplit(file, split = "/"))[4]

  # Get name of the zone as a list
  zone_name <- head(unlist(strsplit(zone_name_path, split = "-")), -2)

  # Collapse name into a single string
  zone_name <- paste(zone_name, collapse = '_')

  # Add name of the zone as a column to the temporary dataset
  df_year_temp$zone <- zone_name

  # Update dataset
  df_continent_avg_temp <- rbind(df_continent_avg_temp, df_year_temp)
}
```

2 Descriptive Statistics

2.1 NASA/GISS

```
# Get dimension of the database
dim(df_temp_anomaly)
```

```
## [1] 141   3
```

```
# Get information about columns
str(df_temp_anomaly)
```

```
## 'data.frame':    141 obs. of  3 variables:
##  $ Year        : int  1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 ...
##  $ No_Smoothing: num  -0.16 -0.08 -0.1 -0.17 -0.28 -0.33 -0.31 -0.36 -0.17 -0.11 ...
##  $ Lowess.5.   : num  -0.09 -0.12 -0.16 -0.2 -0.23 -0.26 -0.27 -0.27 -0.26 -0.25 ...
```

```
# Get summary information about columns
kable(summary(df_temp_anomaly))
```

Year	No_Smoothing	Lowess.5.
Min. :1880	Min. :-0.48000	Min. :-0.41000
1st Qu.:1915	1st Qu.:-0.20000	1st Qu.:-0.22000
Median :1950	Median :-0.07000	Median :-0.04000
Mean :1950	Mean : 0.04858	Mean : 0.04858
3rd Qu.:1985	3rd Qu.: 0.23000	3rd Qu.: 0.22000
Max. :2020	Max. : 1.02000	Max. : 1.01000

```
# Get summary information about columns
kable(summary(df_temp_anomaly))
```

Year	No_Smoothing	Lowess.5.
Min. :1880	Min. :-0.48000	Min. :-0.41000
1st Qu.:1915	1st Qu.:-0.20000	1st Qu.:-0.22000
Median :1950	Median :-0.07000	Median :-0.04000
Mean :1950	Mean : 0.04858	Mean : 0.04858
3rd Qu.:1985	3rd Qu.: 0.23000	3rd Qu.: 0.22000
Max. :2020	Max. : 1.02000	Max. : 1.01000

2.2 Berkley Earth

```
dim(df_continent_avg_temp)
```

```
## [1] 1158    4
```

```
kable(summary(df_continent_avg_temp))
```

year	yearly_anomaly	lowess	zone
Min. :1750	Min. :-1.684333	Min. :-0.769483	Length:1158
1st Qu.:1887	1st Qu.:-0.402314	1st Qu.:-0.369790	Class :character
Median :1935	Median :-0.096292	Median :-0.120499	Mode :character
Mean :1928	Mean :-0.009384	Mean :-0.006192	NA
3rd Qu.:1979	3rd Qu.: 0.337562	3rd Qu.: 0.176794	NA
Max. :2020	Max. : 2.163750	Max. : 1.750654	NA

```
str(df_continent_avg_temp)
```

```
## 'data.frame':    1158 obs. of  4 variables:
##  $ year          : num  1880 1881 1882 1883 1884 ...
##  $ yearly_anomaly: num  -0.282 -0.206 -0.487 -0.497 -0.767 ...
##  $ lowess        : num  -0.498 -0.49 -0.481 -0.472 -0.462 ...
##  $ zone          : chr  "africa" "africa" "africa" "africa" ...
```

```
kable(head(df_continent_avg_temp))
```

year	yearly_anomaly	lowess	zone
1880	-0.2818000	-0.4982481	africa
1881	-0.2058000	-0.4901990	africa
1882	-0.4874167	-0.4814832	africa
1883	-0.4966364	-0.4721923	africa
1884	-0.7671429	-0.4624419	africa
1885	-0.4330000	-0.4523444	africa

```
kable(describe(df_continent_avg_temp)[1:7])
```

	vars	n	mean	sd	median	trimmed	mad
year	1	1158	1927.5077720	62.6804564	1935.0000000	1932.4202586	68.1996000
yearly_anomaly	2	1158	-0.0093837	0.6044417	-0.0962917	-0.0401278	0.5353421
lowess	3	1158	-0.0061921	0.4912716	-0.1204991	-0.0676041	0.4048334
zone*	4	1158	4.1908463	1.8243581	4.0000000	4.2381466	1.4826000

```
kable(describe(df_continent_avg_temp)[7:13])
```

	mad	min	max	range	skew	kurtosis	se
year	68.1996000	1750.0000000	2020.000000	270.000000	-0.6111347	-0.2285290	1.8419501
yearly_anomaly	0.5353421	-1.6843333	2.163750	3.848083	0.4930823	0.3923491	0.0177623
lowess	0.4048334	-0.7694827	1.750654	2.520136	1.1026295	0.7690255	0.0144367
zone*	1.4826000	1.0000000	7.000000	6.000000	-0.1638085	-0.8302282	0.0536112

2.2.1 Descriptitive statistics per continent

Using describeBy() function of psych package.

```
describeBy(df_continent_avg_temp, df_continent_avg_temp$zone)
```

```
##
##  Descriptive statistics by group
## group: africa
##                vars   n    mean    sd  median trimmed   mad     min     max
## year              1 141 1950.00 40.85 1950.00 1950.00 51.89 1880.00 2020.00
## yearly_anomaly    2 141    0.11  0.49    0.02    0.07  0.41   -0.81    1.44
## lowess            3 141    0.10  0.44   -0.02    0.05  0.44   -0.50    1.15
## zone*             4 141    1.00  0.00    1.00    1.00  0.00    1.00    1.00
##                 range skew kurtosis   se
## year           140.00 0.00    -1.23 3.44
## yearly_anomaly   2.25 0.68    -0.21 0.04
## lowess           1.65 0.88    -0.29 0.04
## zone*            0.00  NaN      NaN 0.00
## ------------------------------------------------------------
## group: antarctica
##                vars  n    mean    sd  median trimmed   mad     min     max
## year              1 65 1988.00 18.91 1988.00 1988.00 23.72 1956.00 2020.00
## yearly_anomaly    2 65    0.36  0.55    0.41    0.37  0.48   -0.96    1.41
## lowess            3 65    0.35  0.36    0.32    0.36  0.29   -0.39    1.16
## zone*             4 65    1.00  0.00    1.00    1.00  0.00    1.00    1.00
##                range  skew kurtosis   se
## year           64.00  0.00    -1.26 2.35
## yearly_anomaly  2.37 -0.12    -0.36 0.07
## lowess          1.55  0.01    -0.57 0.04
## zone*           0.00   NaN      NaN 0.00
## ------------------------------------------------------------
## group: asia
##                vars   n    mean    sd  median trimmed   mad     min     max
## year              1 182 1929.50 52.68 1929.50 1929.50 67.46 1839.00 2020.00
## yearly_anomaly    2 182    0.00  0.61   -0.14   -0.06  0.44   -1.55    2.11
## lowess            3 182    0.01  0.55   -0.11   -0.07  0.34   -0.77    1.68
## zone*             4 182    1.00  0.00    1.00    1.00  0.00    1.00    1.00
##                 range skew kurtosis   se
## year           181.00 0.00    -1.22 3.91
## yearly_anomaly   3.66 0.92     1.00 0.05
## lowess           2.45 1.35     1.18 0.04
## zone*            0.00  NaN      NaN 0.00
## ------------------------------------------------------------
## group: europe
##                vars   n    mean    sd  median trimmed    mad     min     max
## year              1 271 1885.00 78.38 1885.00 1885.00 100.82 1750.00 2020.00
## yearly_anomaly    2 271   -0.12  0.69   -0.15   -0.16   0.63   -1.63    2.16
## lowess            3 271   -0.09  0.51   -0.31   -0.19   0.23   -0.62    1.75
## zone*             4 271    1.00  0.00    1.00    1.00   0.00    1.00    1.00
##                 range skew kurtosis   se
## year           270.00 0.00    -1.21 4.76
## yearly_anomaly   3.79 0.56     0.33 0.04
## lowess           2.37 1.77     2.72 0.03
```

```
## zone*            0.00  NaN      NaN 0.00
## ------------------------------------------------------------
## group: north_america
##                vars   n    mean    sd  median trimmed   mad     min     max
## year              1 198 1921.50 57.30 1921.50 1921.50 73.39 1823.00 2020.00
## yearly_anomaly    2 198   -0.08  0.65   -0.17   -0.12  0.62   -1.68    1.94
## lowess            3 198   -0.10  0.50   -0.18   -0.17  0.50   -0.66    1.26
## zone*             4 198    1.00  0.00    1.00    1.00  0.00    1.00    1.00
##                 range skew kurtosis   se
## year           197.00 0.00    -1.22 4.07
## yearly_anomaly   3.63 0.59     0.25 0.05
## lowess           1.93 1.00     0.26 0.04
## zone*            0.00  NaN      NaN 0.00
## ------------------------------------------------------------
## group: oceania
##                vars   n    mean    sd  median trimmed   mad     min     max
## year              1 145 1948.00 42.00 1948.00 1948.00 53.37 1876.00 2020.00
## yearly_anomaly    2 145    0.06  0.48   -0.04    0.03  0.47   -0.88    1.42
## lowess            3 145    0.06  0.39   -0.16    0.00  0.15   -0.28    1.15
## zone*             4 145    1.00  0.00    1.00    1.00  0.00    1.00    1.00
##                 range skew kurtosis   se
## year           144.00 0.00    -1.22 3.49
## yearly_anomaly   2.29 0.60    -0.17 0.04
## lowess           1.44 1.19     0.19 0.03
## zone*            0.00  NaN      NaN 0.00
## ------------------------------------------------------------
## group: south_america
##                vars   n    mean    sd  median trimmed   mad     min     max
## year              1 156 1942.07 45.91 1942.50 1942.50 57.82 1856.00 2020.00
## yearly_anomaly    2 156   -0.07  0.50   -0.13   -0.09  0.48   -1.60    1.24
## lowess            3 156   -0.08  0.45   -0.15   -0.12  0.40   -0.73    1.05
## zone*             4 156    1.00  0.00    1.00    1.00  0.00    1.00    1.00
##                 range  skew kurtosis   se
## year           164.00 -0.05    -1.15 3.68
## yearly_anomaly   2.84  0.32    -0.04 0.04
## lowess           1.78  0.73    -0.43 0.04
## zone*            0.00   NaN      NaN 0.00
```

Manual approach.

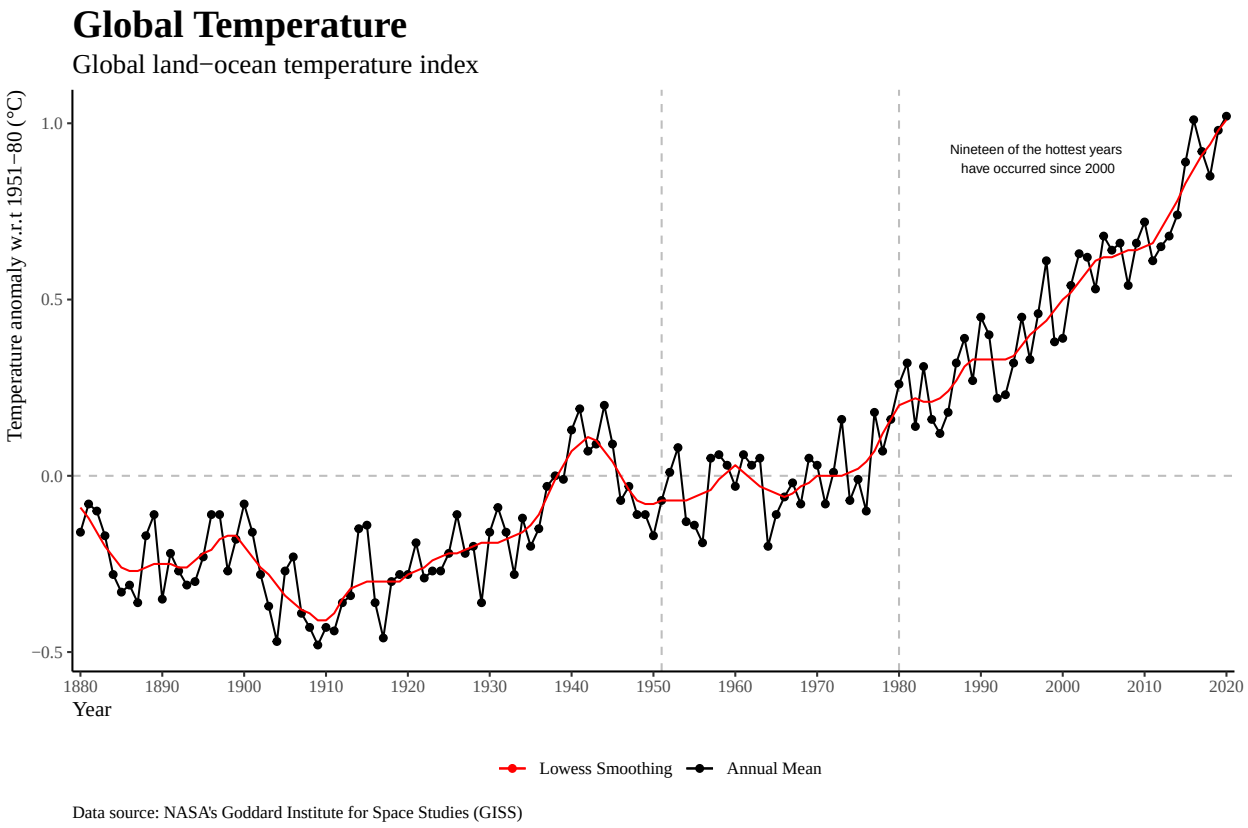
```
# Name of continents
continents <- c("africa", "europe", "asia", "oceania",
                "north_america", "south_america", "antarctica")

# Create empty vectors
min_year <- c()
min_temp_anom <- c()
max_temp_anom <- c()
mean_temp_anom <- c()
var_temp_anom <- c()

# Loop through all the continents
for (continent in continents){
```

```
  # Filter the continent
  df_continent <- df_continent_avg_temp[df_continent_avg_temp$zone == continent, ]

  # Get descriptive statistics per continent
  min_year <- c(min_year, min(df_continent$year))
  min_temp_anom <- c(min_temp_anom, min(df_continent$yearly_anomaly))
  max_temp_anom <- c(max_temp_anom, max(df_continent$yearly_anomaly))
  mean_temp_anom <- c(mean_temp_anom, mean(df_continent$yearly_anomaly))
  var_temp_anom <- c(var_temp_anom, var(df_continent$yearly_anomaly))
}

df_summary_continent <- data.frame(continents, min_year, min_temp_anom,
                                   max_temp_anom, mean_temp_anom, var_temp_anom)

kable(df_summary_continent)
```

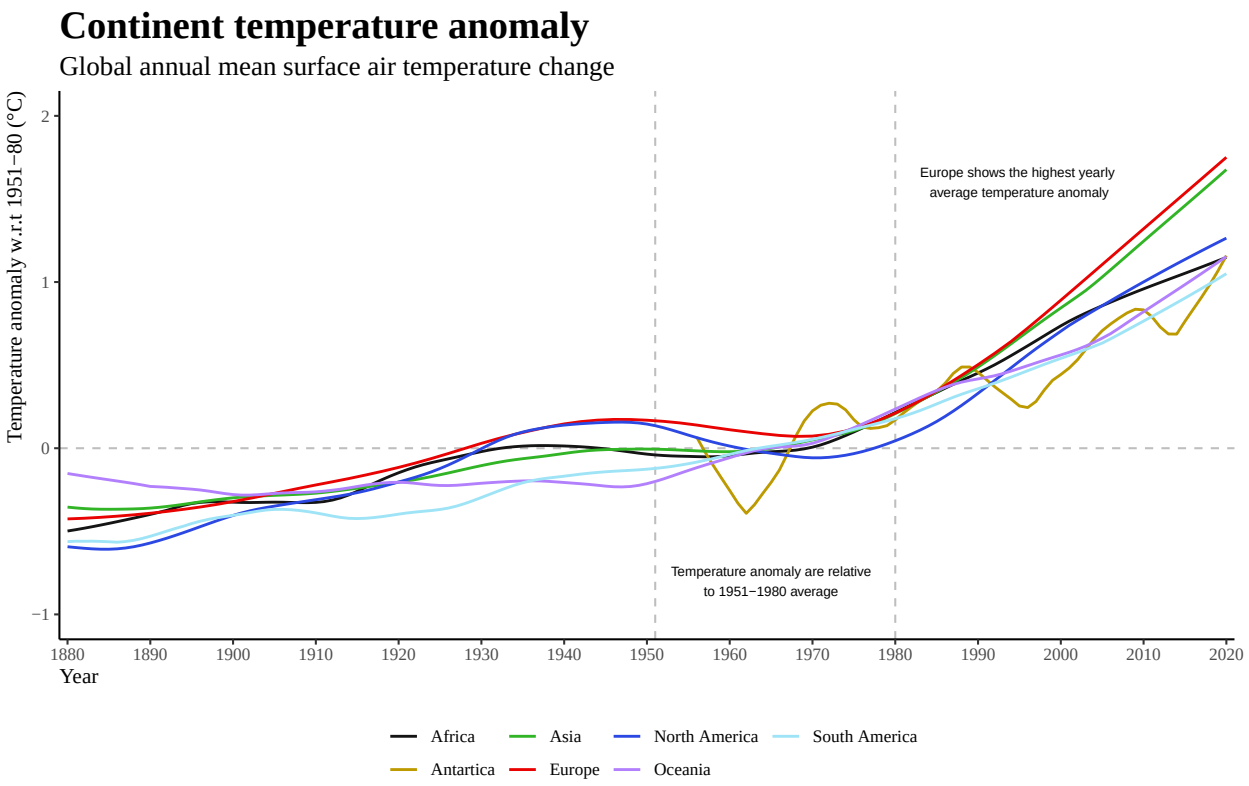
continents	min_year	min_temp_anom	max_temp_anom	mean_temp_anom	var_temp_anom
africa	1880	-0.8090000	1.443167	0.1101983	0.2410926
europe	1750	-1.6291667	2.163750	-0.1166967	0.4793055
asia	1839	-1.5506000	2.111417	0.0031615	0.3752881
oceania	1876	-0.8770833	1.416000	0.0612455	0.2329804
north_america	1823	-1.6843333	1.943500	-0.0828213	0.4250174
south_america	1856	-1.6050000	1.237333	-0.0720385	0.2505224
antarctica	1956	-0.9570000	1.409417	0.3600167	0.2979611

3 Data Visualization

3.1 Global temperature



3.2 Temperature anomaly trend for each continent



Data source: Berkley Earth.

3.3 Temperature records months and regions

```
# Order data by yearly anomaly
df_top_year <- df_continent_avg_temp[order(-df_continent_avg_temp$yearly_anomaly), ]
kable(head(df_top_year, 10))
```

	year	yearly_anomaly	lowess	zone
659	2020	2.163750	1.750654	europe
388	2020	2.111417	1.676391	asia
853	2016	1.943500	1.162198	north_america
658	2019	1.846000	1.707456	europe
653	2014	1.817583	1.492195	europe
654	2015	1.802917	1.535166	europe
657	2018	1.757583	1.664258	europe
847	2010	1.729000	1.000779	north_america
387	2019	1.688917	1.632638	asia
385	2017	1.685833	1.545925	asia

```
df_eu <- df_continent_avg_temp[df_continent_avg_temp$zone == 'europe', ]
write.csv(df_eu, "data/Berkley Earth/df_yearly_anomaly_EU.csv", row.names = TRUE)
```