

Guide 3: Analyzing & Visual Display of Climate Trends

Marc Los Huertos

February 3, 2025 (ver. 0.73)

Contents

1	Introduction	2
1.1	Goals	2
1.2	Weather vs. Climate	2
1.3	Approach	3
1.4	R Code with Custom Functions	3
1.5	Before Starting the Process	4
1.5.1	Reading csv and loading R.data	4
2	Analyzing and Visualizing Monthly Trends	4
2.1	Linear Regression: TMAX, TMIN, and PRCP YEAR, by=YEAR	4
2.2	Plotting a Simple Trend Line	5
2.3	Marc's Custom Function to Plot a Station Trend	7
2.4	Accelerataion of Trend in 1975	10
2.5	Filtering Seasonal Effect	15
2.5.1	Method 1: Filtering by Monthly Mean	16
2.5.2	Method 2: Polynomial Filter	16
3	Extreme Events—Using Daily Records	16
3.1	Complicated Nature of Rainfall Patterns	16
3.2	Drought	16
3.3	Standardized Precipitation Evapotranspiration Index (SPEI)	17
3.3.1	Days in a Row without Rain	17
3.3.2	Changes in the Rainfall Probability Distributions	18
3.4	Record Setting Temperature Records	19
3.5	KISS	20
3.5.1	Change Point Analysis	20
3.6	Temp & Precipitation Probability	20
3.7	Using library densEstBayes	20
3.8	Probability Distributions	20
3.9	Evaluating Records	20
3.10	Export Options	20

4	Sea Surface Temperature Data – SURP PROJECT WAITING TO HAPPEN	21
5	Satellite Data	21
6	Ice-Core Data	21
7	Marc’s Custom Functions	21
7.1	monthlyTrend.fun	21
7.1.1	Deprecated Loop Code	22
8	Troubleshooting	23
8.1	Loading and Reading Data into the Knit Environment	23
9	Mapping (GIS) and Weather Stations	24
9.1	Simple Mapping of All US and Canada Weather Stations	24
9.2	Map US Weather Stations	24
9.3	Simple Mapping of Selected State’s 15 oldest, active Weather Stations	26
9.4	More Complex Mapping of Stations	27
10	Mapping Long-term Weather Records	27
10.1	Creating A Basic Trend Plot	27
11	Extreme Events–Using Daily Records	27
12	Conclusions	27

1 Introduction

1.1 Goals

We will use basic linear regression models, i.e. the `lm()` function to determine if there are trends in the data. We will also consider the possibility of non-linear trends, but we will start with the simplest approach. We will evaluate data for extreme events, i.e. the tails of the distribution, and we will also consider the possibility of changes in the distribution over time and records "highs", "lows" and precipitation, and drought.

1.2 Weather vs. Climate

The difference between weather and climate is a measure of time. Weather is what conditions of the atmosphere are over a short period of time, and climate is how the atmosphere "behaves" over relatively long periods of time.

In general, understanding the climate is a question of averages and ranges, of statistical likelihoods, and of repeated or predictable patterns. Weather, on

the other hand, is what we experience on a day-to-day basis, and it might vary wildly from minute to minute, hour to hour, day to day, and season to season.

1.3 Approach

Here's how I might start: Read the EPA summaries and see what stands out as compelling types of analyses. Then read news articles on the region and see what qualifies as "weather" news in the region.

I suggest you consider 3-5 questions that you think are interesting and then we can work together to see if we can answer those questions using the station data.

We need to address several things that might get in the way of our analysis:

1. Determine if there are trends by months
2. Determine if there if trends are more common in recent years
3. Evaluate distribution changes by decade / score
4. Evaluate extreme events

In contrast to Guide 1 and 2, I couldn't separate the explanation of code from the code and statistical analysis. If you have suggestions of how to streamline this, I am all ears!

Moreover, I haven't completed every possible analysis yet, but I tried to post the potential ones that I could think of. If you have some idea that is missing, let me know and we can figure out how to create an analysis for it!

1. Explore various methods to display climate data
2. Use the following [Google Doc](#) to suggest code and/or some functions that can be used to analyze your data. Marc and mentors will meet to see if we can help you with your code.

1.4 R Code with Custom Functions

From the Canvas page, go to the Guide3functions.R file and download the file to your computer. Then upload the file to Rstudio directory you are using for the project. Make sure you have an updated version of the file, this one seems the least stable!

Open the file in Rstudio and run the code, using the "source". button near the top of the editor window.

Run the [Guide3functions.R](#) code and the functions will be loaded into your environment automatically and are designed to help you analyze your station data.

the Guide3functions.R file is living file – so I suggest you check on the updates to the file every few days to see if there have been improvements.

As we get further into the project, these code chunks may require tweaking because of the questions you are interested for your location falls outside the design of the custom functions. Please let Marc or the mentors know to get assistance tweaking the code!

1.5 Before Starting the Process

Before you begin, make sure you have the stations to read into R. Look at the R environment to see that the stations have been loaded in to R.

```
ls()
```

If not, please Slack Marc and mentors, so we help you get these files into the R environment, which might mean running the previous guide again.

1.5.1 Reading csv and loading R.data

If you decide to use the knitting function in R, it turns out that the R environment that the console has is not available for the knitting function. So, I had to create a way to bring the data into the knitted environment. If you would like to do this, see Section [8.1](#).

2 Analyzing and Visualizing Monthly Trends

2.1 Linear Regression: TMAX, TMIN, and PRCP YEAR, by=YEAR

In this section, we'll discuss how we might analyze the trends in the data by month. In other words, is there a trend in January, February, March, etc. Moreover, we might find the trends differ dramatically between months.

Analyze Stations for Trends by Month This function will take the list of dataframes and return a list of linear models for each month. The function will also return a summary of the linear models.

Function: `monthlyTrend.fun()`

Example of how to use the function Here's an example using the list of station dataframes anomalies. Note: I have often spelled the word anomaly wrong!

```
USC00042294.trends <- monthlyTrend.fun(USC00042294.anomalies)
```

Explore Results Table 1 summarizes the monthly trends for TMAX. Admittedly, determining the months with the biggest changes isn't a very good approach for hypothesis testing – it's more like a fishing expedition, but as long as we understand the difference between an a priori hypothesis and an exploratory analysis, we should be okay if we make appropriate conclusions.

We would want to evaluate the trends for TMIN and PRCP too! You can use the Environment tab in Rstudio to see the trend results for all three variables without needing any code, or use/modify this:

```
USC00042294.trends[USC00042294.trends$ELEMENT=="TMAX",]
```

	ELEMENT	MONTH	Estimate	Std. Error	t value	Pr(> t)	r.squared
1	TMAX	1	0.01	0.01	1.44	0.15	0.02
2	TMAX	2	-0.01	0.01	-1.01	0.31	0.01
3	TMAX	3	-0.01	0.01	-1.41	0.16	0.02
4	TMAX	4	-0.00	0.01	-0.44	0.66	0.00
5	TMAX	5	0.00	0.01	0.39	0.70	0.00
6	TMAX	6	0.00	0.00	0.92	0.36	0.01
7	TMAX	7	-0.01	0.00	-1.71	0.09	0.02
8	TMAX	8	-0.01	0.00	-1.49	0.14	0.02
9	TMAX	9	0.01	0.00	2.11	0.04	0.04
10	TMAX	10	0.01	0.01	1.01	0.31	0.01
11	TMAX	11	-0.01	0.01	-2.30	0.02	0.04
12	TMAX	12	-0.00	0.01	-0.37	0.71	0.00

Table 1: TMAX Trends

2.2 Plotting a Simple Trend Line

Here's an example of how to plot a simple trend line for TMIN for January. We are getting ahead of ourselves here, but it's a good example of how to plot a trend line and some of the back and forth we'll need to do between analysis and presentation.

I suggest you select the element/month with the strongest signal. I selected TMIN and January because it had a strong signal), i.e. highly significant with a p-value ≤ 0.05 .

```
par(mar=c(5,5,3,4), las=1, mfrow=c(1,1))

plot(TMIN.a ~ Ymd, data=subset(USC00042294.anomalies$TMIN,
  MONTH==1), las=1, pch=20, cex=.5, col="grey",
  ylab="Temp Anomaly (C) ",
  main="",
```

```

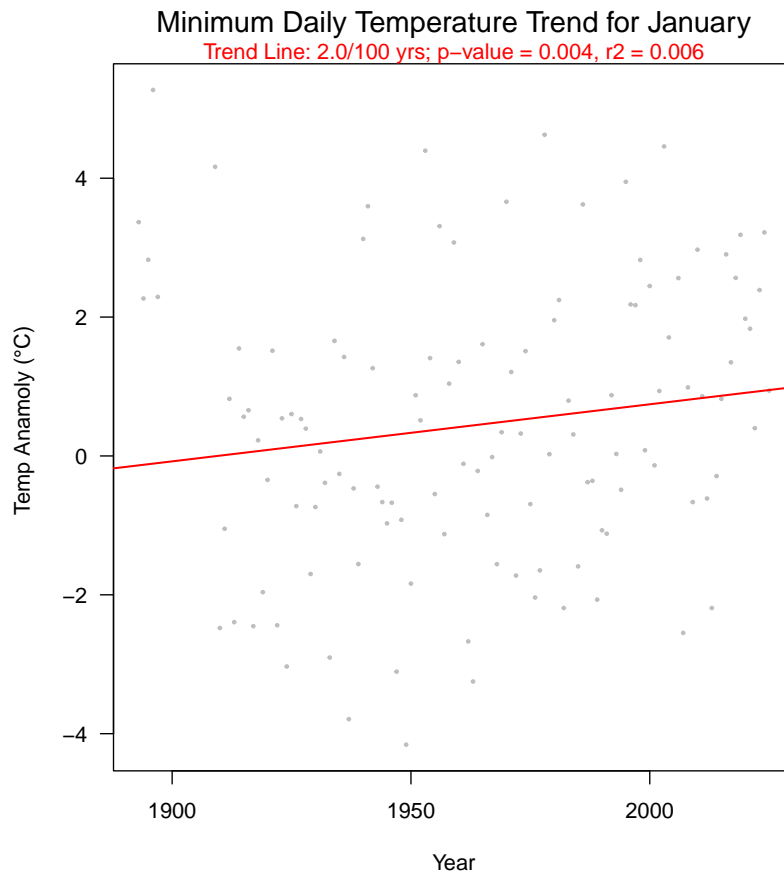
xlab="Year", frame.plot=TRUE)

mtext("Minimum Daily Temperature Trend for January",
      cex=1.3, side=3, line=1)
subtitle <- c("Trend Line: 2.0/100 yrs; p-value = 0.004, r2 = 0.006")
mtext(subtitle, side=3, line=0, cex=1.0, col="red")

USC00042294.lm = lm(TMIN.a ~ Ymd, data=subset(USC00042294.anomalies$TMIN,
      MONTH==1))

abline(coef(USC00042294.lm), col="red", lwd=1.5)

```



2.3 Marc's Custom Function to Plot a Station Trend

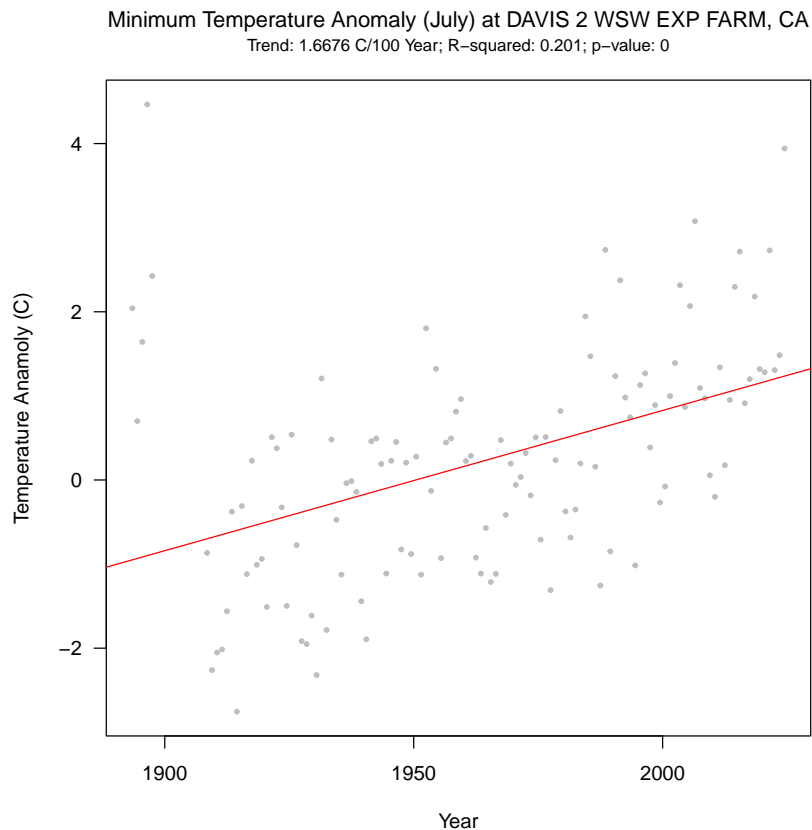
I have a custom function that will plot the trend line for you. It's a bit more complicated than the simple example, but it has a bug that I am still working to fix.

Plot Trends by Month I have created a basic function template, but we'll develop this together when we have a chance. This function will take the list of creates a plot based on three inputs: list of station dataframes, the element, and the month.

Function: `Function: plotTrend.fun()`

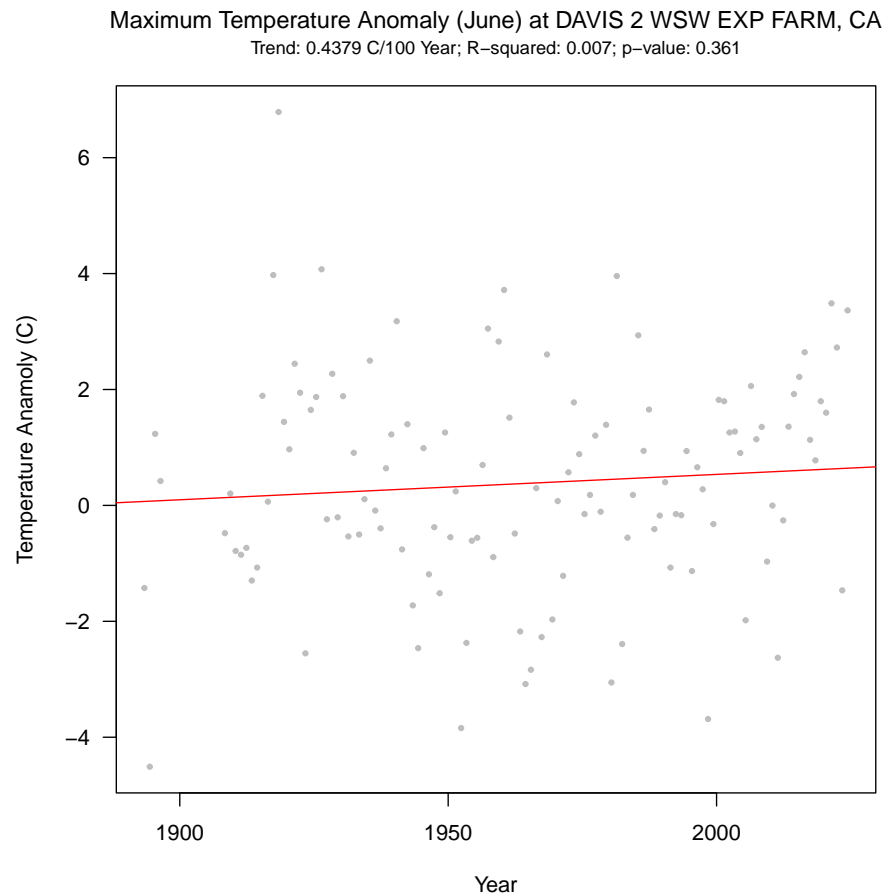
Example of how to use the function Here's an example using the list of station dataframes anomalies. Note: I have often spelled the word anomaly wrong!

```
plotTrend2.fun(USC00042294.anomalies, "TMIN", 7)
```



Explore Results I suggest you look carefully at the plots for hidden trends. In this case, look how the temperature really starts warming after the 1960s. I don't know that that means, but seems important!

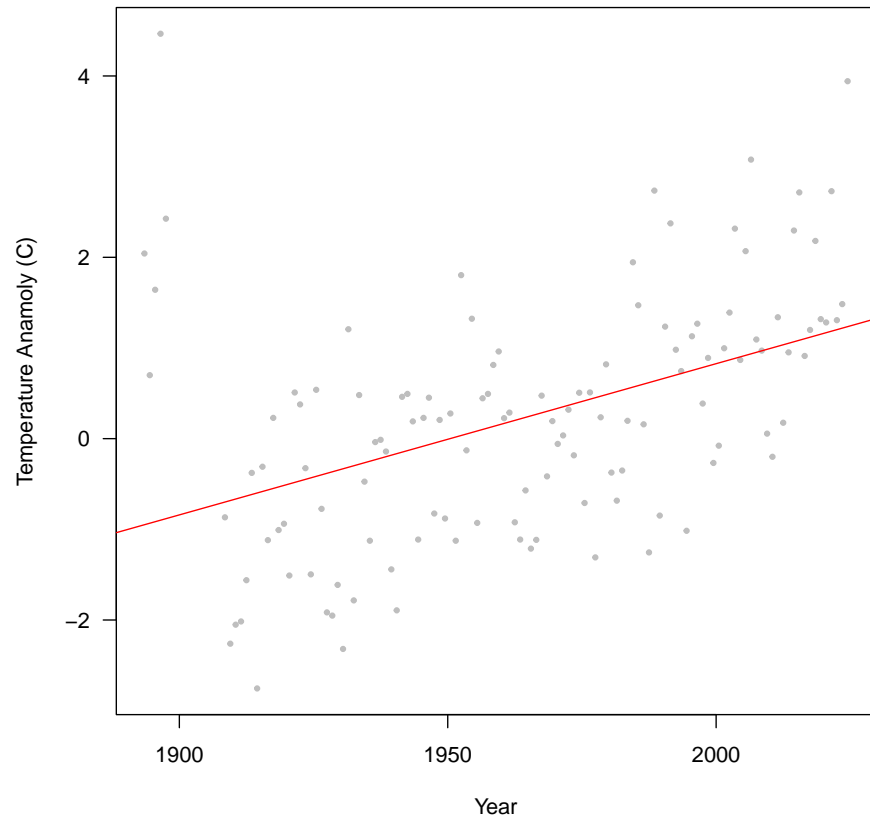
```
plotTrend2.fun(USC00042294.anomalies, "TMAX", 6)
```



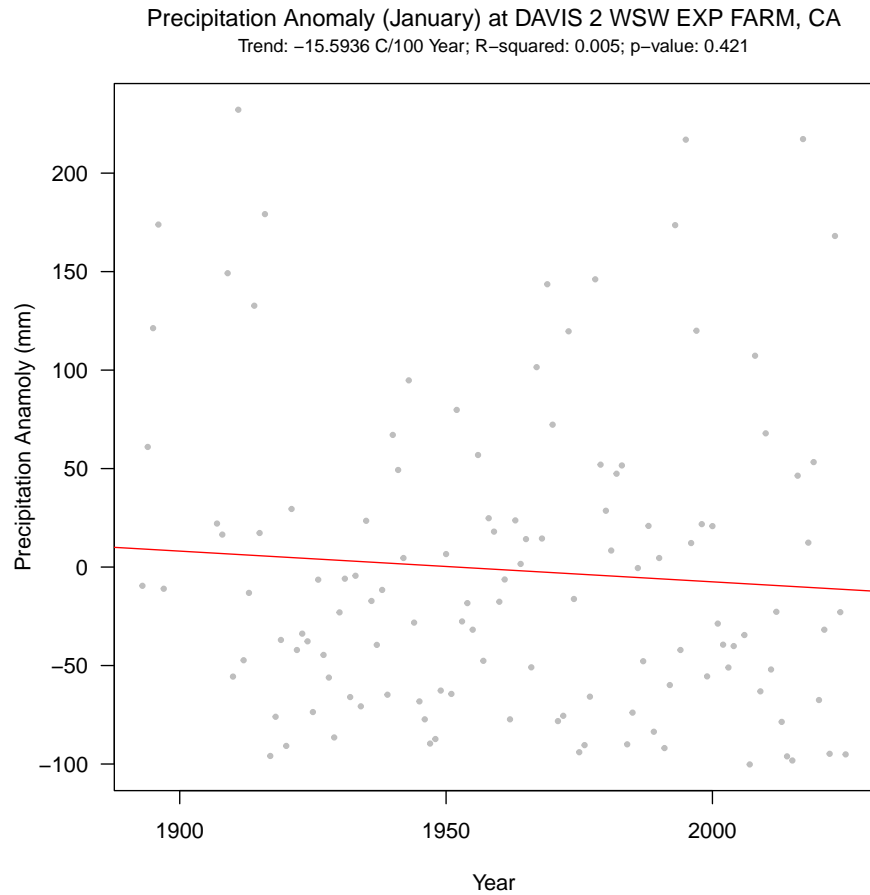
```
plotTrend2.fun(USC00042294.anomalies, "TMIN", 7)
```


Minimum Temperature Anomaly (July) at DAVIS 2 WSW EXP FARM, CA

Trend: 1.6676 C/100 Year; R-squared: 0.201; p-value: 0



```
plotTrend2.fun(USC00042294.anomalies, "PRCP", 1)
```



2.4 Acceleration of Trend in 1975

For this analysis, we compare the slope, p-value, and r-squared between the station's entire record and the record starting in January 1, 1975.

For this we'll plot the entire record, with trend line and then overlay another trend line for the record starting in 1975.

This is code for TMAX ([[1]]) and June. I need to use "predict" to get the trend line for the second model that only includes the data after 1975.

```
USC00042294.lm = lm(TMAX.a ~ Ymd, data=subset(USC00042294.anomalies[[1]], subset=MONTH==6))

USC00042294.post1975 = subset(USC00042294.anomalies[[1]], subset= YEAR > 1974)
USC00042294.post1975.lm = lm(TMAX.a ~ Ymd, data=subset(USC00042294.post1975, subset=MONTH==6))

plot(TMAX.a ~ Ymd, data=subset(USC00042294.anomalies[[1]], subset=MONTH==6), pch=20, cex=.5,
```

```

abline(coef(USC00042294.lm), col="black", lwd=1.5)

points(TMAX.a ~ Ymd, data=subset(USC00042294.post1975, subset=MONTH==6), pch=20, cex=.5, col="red")

new = seq.Date(from=as.Date("1975-01-01"), to=as.Date("2024-12-31"), by="month")
new

##      [1] "1975-01-01" "1975-02-01" "1975-03-01" "1975-04-01" "1975-05-01"
##      [6] "1975-06-01" "1975-07-01" "1975-08-01" "1975-09-01" "1975-10-01"
##     [11] "1975-11-01" "1975-12-01" "1976-01-01" "1976-02-01" "1976-03-01"
##     [16] "1976-04-01" "1976-05-01" "1976-06-01" "1976-07-01" "1976-08-01"
##     [21] "1976-09-01" "1976-10-01" "1976-11-01" "1976-12-01" "1977-01-01"
##     [26] "1977-02-01" "1977-03-01" "1977-04-01" "1977-05-01" "1977-06-01"
##     [31] "1977-07-01" "1977-08-01" "1977-09-01" "1977-10-01" "1977-11-01"
##     [36] "1977-12-01" "1978-01-01" "1978-02-01" "1978-03-01" "1978-04-01"
##     [41] "1978-05-01" "1978-06-01" "1978-07-01" "1978-08-01" "1978-09-01"
##     [46] "1978-10-01" "1978-11-01" "1978-12-01" "1979-01-01" "1979-02-01"
##     [51] "1979-03-01" "1979-04-01" "1979-05-01" "1979-06-01" "1979-07-01"
##     [56] "1979-08-01" "1979-09-01" "1979-10-01" "1979-11-01" "1979-12-01"
##     [61] "1980-01-01" "1980-02-01" "1980-03-01" "1980-04-01" "1980-05-01"
##     [66] "1980-06-01" "1980-07-01" "1980-08-01" "1980-09-01" "1980-10-01"
##     [71] "1980-11-01" "1980-12-01" "1981-01-01" "1981-02-01" "1981-03-01"
##     [76] "1981-04-01" "1981-05-01" "1981-06-01" "1981-07-01" "1981-08-01"
##     [81] "1981-09-01" "1981-10-01" "1981-11-01" "1981-12-01" "1982-01-01"
##     [86] "1982-02-01" "1982-03-01" "1982-04-01" "1982-05-01" "1982-06-01"
##     [91] "1982-07-01" "1982-08-01" "1982-09-01" "1982-10-01" "1982-11-01"
##     [96] "1982-12-01" "1983-01-01" "1983-02-01" "1983-03-01" "1983-04-01"
##    [101] "1983-05-01" "1983-06-01" "1983-07-01" "1983-08-01" "1983-09-01"
##    [106] "1983-10-01" "1983-11-01" "1983-12-01" "1984-01-01" "1984-02-01"
##    [111] "1984-03-01" "1984-04-01" "1984-05-01" "1984-06-01" "1984-07-01"
##    [116] "1984-08-01" "1984-09-01" "1984-10-01" "1984-11-01" "1984-12-01"
##    [121] "1985-01-01" "1985-02-01" "1985-03-01" "1985-04-01" "1985-05-01"
##    [126] "1985-06-01" "1985-07-01" "1985-08-01" "1985-09-01" "1985-10-01"
##    [131] "1985-11-01" "1985-12-01" "1986-01-01" "1986-02-01" "1986-03-01"
##    [136] "1986-04-01" "1986-05-01" "1986-06-01" "1986-07-01" "1986-08-01"
##    [141] "1986-09-01" "1986-10-01" "1986-11-01" "1986-12-01" "1987-01-01"
##    [146] "1987-02-01" "1987-03-01" "1987-04-01" "1987-05-01" "1987-06-01"
##    [151] "1987-07-01" "1987-08-01" "1987-09-01" "1987-10-01" "1987-11-01"
##    [156] "1987-12-01" "1988-01-01" "1988-02-01" "1988-03-01" "1988-04-01"
##    [161] "1988-05-01" "1988-06-01" "1988-07-01" "1988-08-01" "1988-09-01"
##    [166] "1988-10-01" "1988-11-01" "1988-12-01" "1989-01-01" "1989-02-01"
##    [171] "1989-03-01" "1989-04-01" "1989-05-01" "1989-06-01" "1989-07-01"
##    [176] "1989-08-01" "1989-09-01" "1989-10-01" "1989-11-01" "1989-12-01"
##    [181] "1990-01-01" "1990-02-01" "1990-03-01" "1990-04-01" "1990-05-01"
##    [186] "1990-06-01" "1990-07-01" "1990-08-01" "1990-09-01" "1990-10-01"

```

##	[191]	"1990-11-01"	"1990-12-01"	"1991-01-01"	"1991-02-01"	"1991-03-01"
##	[196]	"1991-04-01"	"1991-05-01"	"1991-06-01"	"1991-07-01"	"1991-08-01"
##	[201]	"1991-09-01"	"1991-10-01"	"1991-11-01"	"1991-12-01"	"1992-01-01"
##	[206]	"1992-02-01"	"1992-03-01"	"1992-04-01"	"1992-05-01"	"1992-06-01"
##	[211]	"1992-07-01"	"1992-08-01"	"1992-09-01"	"1992-10-01"	"1992-11-01"
##	[216]	"1992-12-01"	"1993-01-01"	"1993-02-01"	"1993-03-01"	"1993-04-01"
##	[221]	"1993-05-01"	"1993-06-01"	"1993-07-01"	"1993-08-01"	"1993-09-01"
##	[226]	"1993-10-01"	"1993-11-01"	"1993-12-01"	"1994-01-01"	"1994-02-01"
##	[231]	"1994-03-01"	"1994-04-01"	"1994-05-01"	"1994-06-01"	"1994-07-01"
##	[236]	"1994-08-01"	"1994-09-01"	"1994-10-01"	"1994-11-01"	"1994-12-01"
##	[241]	"1995-01-01"	"1995-02-01"	"1995-03-01"	"1995-04-01"	"1995-05-01"
##	[246]	"1995-06-01"	"1995-07-01"	"1995-08-01"	"1995-09-01"	"1995-10-01"
##	[251]	"1995-11-01"	"1995-12-01"	"1996-01-01"	"1996-02-01"	"1996-03-01"
##	[256]	"1996-04-01"	"1996-05-01"	"1996-06-01"	"1996-07-01"	"1996-08-01"
##	[261]	"1996-09-01"	"1996-10-01"	"1996-11-01"	"1996-12-01"	"1997-01-01"
##	[266]	"1997-02-01"	"1997-03-01"	"1997-04-01"	"1997-05-01"	"1997-06-01"
##	[271]	"1997-07-01"	"1997-08-01"	"1997-09-01"	"1997-10-01"	"1997-11-01"
##	[276]	"1997-12-01"	"1998-01-01"	"1998-02-01"	"1998-03-01"	"1998-04-01"
##	[281]	"1998-05-01"	"1998-06-01"	"1998-07-01"	"1998-08-01"	"1998-09-01"
##	[286]	"1998-10-01"	"1998-11-01"	"1998-12-01"	"1999-01-01"	"1999-02-01"
##	[291]	"1999-03-01"	"1999-04-01"	"1999-05-01"	"1999-06-01"	"1999-07-01"
##	[296]	"1999-08-01"	"1999-09-01"	"1999-10-01"	"1999-11-01"	"1999-12-01"
##	[301]	"2000-01-01"	"2000-02-01"	"2000-03-01"	"2000-04-01"	"2000-05-01"
##	[306]	"2000-06-01"	"2000-07-01"	"2000-08-01"	"2000-09-01"	"2000-10-01"
##	[311]	"2000-11-01"	"2000-12-01"	"2001-01-01"	"2001-02-01"	"2001-03-01"
##	[316]	"2001-04-01"	"2001-05-01"	"2001-06-01"	"2001-07-01"	"2001-08-01"
##	[321]	"2001-09-01"	"2001-10-01"	"2001-11-01"	"2001-12-01"	"2002-01-01"
##	[326]	"2002-02-01"	"2002-03-01"	"2002-04-01"	"2002-05-01"	"2002-06-01"
##	[331]	"2002-07-01"	"2002-08-01"	"2002-09-01"	"2002-10-01"	"2002-11-01"
##	[336]	"2002-12-01"	"2003-01-01"	"2003-02-01"	"2003-03-01"	"2003-04-01"
##	[341]	"2003-05-01"	"2003-06-01"	"2003-07-01"	"2003-08-01"	"2003-09-01"
##	[346]	"2003-10-01"	"2003-11-01"	"2003-12-01"	"2004-01-01"	"2004-02-01"
##	[351]	"2004-03-01"	"2004-04-01"	"2004-05-01"	"2004-06-01"	"2004-07-01"
##	[356]	"2004-08-01"	"2004-09-01"	"2004-10-01"	"2004-11-01"	"2004-12-01"
##	[361]	"2005-01-01"	"2005-02-01"	"2005-03-01"	"2005-04-01"	"2005-05-01"
##	[366]	"2005-06-01"	"2005-07-01"	"2005-08-01"	"2005-09-01"	"2005-10-01"
##	[371]	"2005-11-01"	"2005-12-01"	"2006-01-01"	"2006-02-01"	"2006-03-01"
##	[376]	"2006-04-01"	"2006-05-01"	"2006-06-01"	"2006-07-01"	"2006-08-01"
##	[381]	"2006-09-01"	"2006-10-01"	"2006-11-01"	"2006-12-01"	"2007-01-01"
##	[386]	"2007-02-01"	"2007-03-01"	"2007-04-01"	"2007-05-01"	"2007-06-01"
##	[391]	"2007-07-01"	"2007-08-01"	"2007-09-01"	"2007-10-01"	"2007-11-01"
##	[396]	"2007-12-01"	"2008-01-01"	"2008-02-01"	"2008-03-01"	"2008-04-01"
##	[401]	"2008-05-01"	"2008-06-01"	"2008-07-01"	"2008-08-01"	"2008-09-01"
##	[406]	"2008-10-01"	"2008-11-01"	"2008-12-01"	"2009-01-01"	"2009-02-01"
##	[411]	"2009-03-01"	"2009-04-01"	"2009-05-01"	"2009-06-01"	"2009-07-01"

```
## [416] "2009-08-01" "2009-09-01" "2009-10-01" "2009-11-01" "2009-12-01"
## [421] "2010-01-01" "2010-02-01" "2010-03-01" "2010-04-01" "2010-05-01"
## [426] "2010-06-01" "2010-07-01" "2010-08-01" "2010-09-01" "2010-10-01"
## [431] "2010-11-01" "2010-12-01" "2011-01-01" "2011-02-01" "2011-03-01"
## [436] "2011-04-01" "2011-05-01" "2011-06-01" "2011-07-01" "2011-08-01"
## [441] "2011-09-01" "2011-10-01" "2011-11-01" "2011-12-01" "2012-01-01"
## [446] "2012-02-01" "2012-03-01" "2012-04-01" "2012-05-01" "2012-06-01"
## [451] "2012-07-01" "2012-08-01" "2012-09-01" "2012-10-01" "2012-11-01"
## [456] "2012-12-01" "2013-01-01" "2013-02-01" "2013-03-01" "2013-04-01"
## [461] "2013-05-01" "2013-06-01" "2013-07-01" "2013-08-01" "2013-09-01"
## [466] "2013-10-01" "2013-11-01" "2013-12-01" "2014-01-01" "2014-02-01"
## [471] "2014-03-01" "2014-04-01" "2014-05-01" "2014-06-01" "2014-07-01"
## [476] "2014-08-01" "2014-09-01" "2014-10-01" "2014-11-01" "2014-12-01"
## [481] "2015-01-01" "2015-02-01" "2015-03-01" "2015-04-01" "2015-05-01"
## [486] "2015-06-01" "2015-07-01" "2015-08-01" "2015-09-01" "2015-10-01"
## [491] "2015-11-01" "2015-12-01" "2016-01-01" "2016-02-01" "2016-03-01"
## [496] "2016-04-01" "2016-05-01" "2016-06-01" "2016-07-01" "2016-08-01"
## [501] "2016-09-01" "2016-10-01" "2016-11-01" "2016-12-01" "2017-01-01"
## [506] "2017-02-01" "2017-03-01" "2017-04-01" "2017-05-01" "2017-06-01"
## [511] "2017-07-01" "2017-08-01" "2017-09-01" "2017-10-01" "2017-11-01"
## [516] "2017-12-01" "2018-01-01" "2018-02-01" "2018-03-01" "2018-04-01"
## [521] "2018-05-01" "2018-06-01" "2018-07-01" "2018-08-01" "2018-09-01"
## [526] "2018-10-01" "2018-11-01" "2018-12-01" "2019-01-01" "2019-02-01"
## [531] "2019-03-01" "2019-04-01" "2019-05-01" "2019-06-01" "2019-07-01"
## [536] "2019-08-01" "2019-09-01" "2019-10-01" "2019-11-01" "2019-12-01"
## [541] "2020-01-01" "2020-02-01" "2020-03-01" "2020-04-01" "2020-05-01"
## [546] "2020-06-01" "2020-07-01" "2020-08-01" "2020-09-01" "2020-10-01"
## [551] "2020-11-01" "2020-12-01" "2021-01-01" "2021-02-01" "2021-03-01"
## [556] "2021-04-01" "2021-05-01" "2021-06-01" "2021-07-01" "2021-08-01"
## [561] "2021-09-01" "2021-10-01" "2021-11-01" "2021-12-01" "2022-01-01"
## [566] "2022-02-01" "2022-03-01" "2022-04-01" "2022-05-01" "2022-06-01"
## [571] "2022-07-01" "2022-08-01" "2022-09-01" "2022-10-01" "2022-11-01"
## [576] "2022-12-01" "2023-01-01" "2023-02-01" "2023-03-01" "2023-04-01"
## [581] "2023-05-01" "2023-06-01" "2023-07-01" "2023-08-01" "2023-09-01"
## [586] "2023-10-01" "2023-11-01" "2023-12-01" "2024-01-01" "2024-02-01"
## [591] "2024-03-01" "2024-04-01" "2024-05-01" "2024-06-01" "2024-07-01"
## [596] "2024-08-01" "2024-09-01" "2024-10-01" "2024-11-01" "2024-12-01"
```

```
predict.lm(USC00042294.post1975.lm, interval="confidence")
```

```
##           fit           lwr          upr
## 606 0.840244190 0.30573278 1.3747556
## 607 0.871638586 0.32025490 1.4230223
## 610 0.903032983 0.33348959 1.4725764
## 613 0.965907788 0.35659397 1.5752216
## 614 0.934427380 0.34555595 1.5232988
```

```

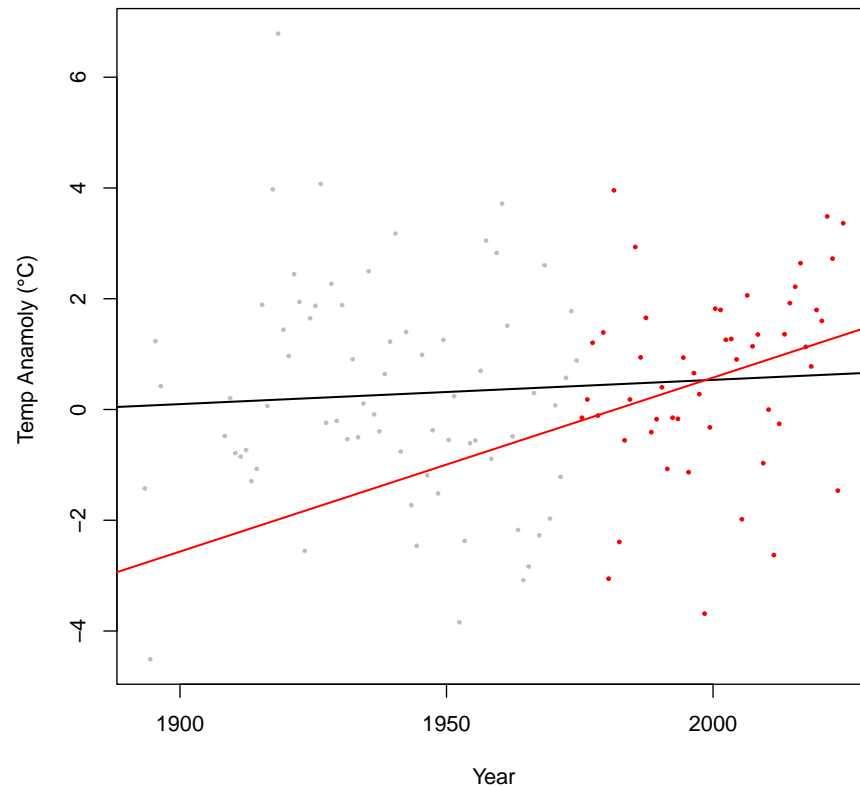
## 619 0.431772986 -0.05065110 0.9141971
## 620 0.777369385 0.27226972 1.2824690
## 621 0.997302185 0.36664607 1.6279583
## 623 0.808763781 0.28975196 1.3277756
## 627 0.463253395 -0.01063310 0.9371399
## 631 -0.165064596 -1.04512541 0.7149962
## 633 1.028696581 0.37583739 1.6815558
## 635 0.494647792 0.02724744 0.9620481
## 639 -0.133670200 -0.98671786 0.7193775
## 642 0.620311390 0.15728660 1.0833362
## 644 0.557436585 0.09661668 1.0182565
## 645 1.060090978 0.38425277 1.7359292
## 646 0.745974988 0.25311370 1.2388363
## 647 0.526042188 0.06301740 0.9890670
## 653 1.185754576 0.41158635 1.9599228
## 654 0.651705786 0.18430543 1.1191061
## 655 -0.196545005 -1.10401102 0.7109210
## 657 1.091571386 0.39198891 1.7911539
## 659 0.023387795 -0.70050566 0.7472812
## 665 1.217234985 0.41713537 2.0173346
## 666 0.683100183 0.20921369 1.1569867
## 667 0.400378590 -0.09248269 0.8932399
## 669 1.122965783 0.39907233 1.8468592
## 670 0.588916993 0.12809709 1.0497369
## 671 0.054782192 -0.64480028 0.7543647
## 677 1.248629381 0.42224705 2.0750117
## 679 0.180445790 -0.42886803 0.7897596
## 681 1.154360180 0.40558652 1.9031338
## 682 0.086262600 -0.58957561 0.7621008
## 689 1.280023778 0.42697612 2.1330714
## 691 -0.102275803 -0.92865814 0.7241065
## 693 0.149051393 -0.48160472 0.7797075
## 694 0.117656997 -0.53520219 0.7705162
## 698 0.368984193 -0.13611547 0.8740839
## 699 0.337589797 -0.18142202 0.8566016
## 701 1.311418174 0.43135736 2.1914790
## 703 -0.070881407 -0.87098102 0.7292182
## 709 1.342898583 0.43543257 2.2503646
## 714 0.714580592 0.23215650 1.1970047
## 715 -0.039400998 -0.81356922 0.7347672
## 719 -0.008006602 -0.75678026 0.7407671
## 721 0.243320595 -0.32622280 0.8128640
## 722 0.211926198 -0.37694523 0.8007976
## 723 0.306109388 -0.22840202 0.8406208
## 724 0.274714992 -0.27666870 0.8260987

```

```
lines(USC00042294.post1975$Ymd, USC00042294.post1975.lm$fitted.values, col="red", lwd=1.5)

## Error in xy.coords(x, y): 'x' and 'y' lengths differ

abline(coef(USC00042294.post1975.lm), col="red", lwd=1.5)
```



2.5 Filtering Seasonal Effect

Since we are looking at trends over years, it might be useful to filter out the seasonal effects.

There are several ways to filter out seasonal effects. The easiest way is subtract the mean value for each date, but that's tricky because every four years there is an extra day in February – although there are ways to deal with this, a more straight forward way is to use mean monthly values to capture the seasonality for each month. With 12 months, this is a pretty good approach

because the pretty good resolution is pretty good when the station has complete records.

2.5.1 Method 1: Filtering by Monthly Mean

One way of doing this is creating a matrix of values for each month and then subtracting the mean value for each month, for the whole record, not just the 1960-1990 “normals”. If this is of interest, we can help you with this.

2.5.2 Method 2: Polynomial Filter

Another method is to use a polynomial filter. Perhaps, someday, I’ll work on this. Great student project to be followed up with.

```
# fit polynomial:  $x^2b_1 + x^2b_2 + \dots + b_n$ 

# create time series object
#X = [i%365 for i in range(0, len(series))]
# y = series.values

# degree = 4
#coef = polyfit(X, y, degree)
# print('Coefficients: %s' % coef)
# create curve
```

3 Extreme Events—Using Daily Records

3.1 Complicated Nature of Rainfall Patterns

Rainfall trends are tough. Extreme events can occur in 24 hours or over long periods that might result in floods or droughts. Each region might have different patterns, so developing a consistent approach is tough.

We can look for trends in monthly averages, number of days without rain (important in tropics), and/or extreme events based on daily or hourly data.

In addition, the definition of extreme events is highly regionally specific. Thus, if this is of interest, let us know and we’ll help you develop code!

Rainfall totals by season might be a useful way to think about changes, because the rainfall is often seasonal, I wonder if we can see patterns by season.

3.2 Drought

Days without rain...within a calendar year... bleed over between years isn’t captured.. This is screwed up, Drought.run needs work.

3.3 Standardized Precipitation Evapotranspiration Index (SPEI)

The Standardized Precipitation Evapotranspiration Index (SPEI) is an extension of the widely used Standardized Precipitation Index (SPI). The SPEI is designed to take into account both precipitation and potential evapotranspiration (PET) in determining drought. Thus, unlike the SPI, the SPEI captures the main impact of increased temperatures on water demand.

I don't know if we can do this yet, but I am working on the code to develop the SPEI using a publish R package <https://cran.r-project.org/web/packages/SPEI/>.

3.3.1 Days in a Row without Rain

One proxy for drought is a long period without rain. This is a bit tricky because it's not just the number of days without rain, but the number of days without rain in a row. So, I found a function call `rle()` that counts the number of days in a row without rain.

```
## function (station = USC00040693, threshold = 0.1)
## {
##   drought.df <- subset(station, ELEMENT == "PRCP")
##   x <- drought.df$VALUE
##   length(x)
##   drought = x < threshold
##   str(drought)
##   drought = rle(drought)
##   str(drought)
##   rle.values <- as.data.frame(do.call(cbind, drought))
##   str(rle.values)
##   drought.df$Drought = as.logical(rep(rle.values$values, times = rle.values$lengths))
##   head(drought.df)
##   drought.df$Length = rep(rle.values$lengths, rle.values$lengths)
##   head(drought.df)
##   if ("Ymd" %in% names(drought.df) == FALSE) {
##     drought.df$Ymd = as.Date(as.character(drought.df$DATE),
##                               format = "%Y%m%d")
##     drought.df$MONTH = as.numeric(format(drought.df$Ymd,
##                                           "%m"))
##     drought.df$YEAR = as.numeric(format(drought.df$Ymd, "%Y"))
##   }
##   drought.df$WY = ifelse(drought.df$MONTH < 10, drought.df$YEAR,
##                           drought.df$YEAR + 1)
##   str(drought.df)
##   aggregate(drought.df$Drought, by = list(YEAR = drought.df$YEAR,
##                                           MONTH = drought.df$MONTH), FUN = sum)
```

```
## DroughtperYear = aggregate(drought.df$Drought, by = list(WY = drought.df$WY),
## FUN = sum)
## RecordperYear = aggregate(drought.df$DATE, by = list(WY = drought.df$WY),
## FUN = length)
## drought = merge(RecordperYear, DroughtperYear, by = "WY")
## head(drought)
## drought$DroughtPerYear = round(drought$x.y/drought$x.x *
## 100)
## head(drought)
## return(drought)
## }
```

Example of how to use the function Here's an example using the station dataframe.

```
USC00040693.drought <- droughtCount.fun(USC00040693, threshold=0.1)
```

3.3.2 Changes in the Rainfall Probability Distributions

In this example, we could either use `rnorm` or a kernel density estimate, for each 10 years or 20 years segments and see if there is a difference over time. In general, these can be pretty compelling. But it's the mean that is really driving the analysis, so that might be something to look at by decade to see if there is a trend.

	Decade	Fall	Spring	Summer	Winter
1	1890.00	54.02	65.02	4.06	114.48
2	1900.00	32.56	60.62	2.86	127.27
3	1910.00	27.54	36.74	2.44	127.82
4	1920.00	34.28	38.41	2.16	92.32
5	1930.00	21.94	43.33	2.38	97.58
6	1940.00	33.71	54.42	1.59	105.02
7	1950.00	32.93	46.19	2.16	122.48
8	1960.00	40.64	44.20	3.00	107.50
9	1970.00	50.57	44.26	3.61	103.29
10	1980.00	49.78	54.81	1.82	112.22
11	1990.00	39.47	48.95	3.22	138.97
12	2000.00	33.30	47.23	1.33	136.12
13	2010.00	31.17	65.72	2.96	109.66
14	2020.00	0.83	21.07	0.22	28.06

3.4 Record Setting Temperature Records

In many cases, people seem to "feel" how temperature has been changing over time, and new records seem to capture the attention in the media. So, we'll create a updated record of maximum temperatures and display them.

How to do this? These might be the steps!

1. First, we'll calculate the mean temperature for the entire period.
2. We might sort the data set, by highest/lowest temperature for each day of the year.
3. Then we can see if the frequency of the highest/lowest temperatures is increasing over time.
4. We'll then create a new column that will be the minimum temperature for each day.
5. We'll then create a new column that will be the maximum temperature for each day, but only if it is the maximum temperature for that day.
6. We'll then create a new column that will be the minimum temperature for each day, but only if it is the minimum temperature for that day.

This is a common way to communicate temperatures changes. I suspect we have a better sense of change when we notice "extreme" events, and this also fits the news media's need for "new" stories.

```
ggplot( ) +  
  geom_bar(data = records, aes(x=Year, y=Num, fill=Group),  
    stat="identity", position="identity") +  
  xlim(min(CHCND$Year), max(CHCND$Year)-1) +  
  ylab("Number of Extreme Temps") + # for the y axis label  
  scale_fill_manual("Legend",  
    values = c("Record Highs" = "red", "Record Lows" = "blue"))
```

```
#TMIN in the March, at station USC00042294  
par(mfrow=c(1,1))  
station = subset(USC00042294.anomalies[[2]], subset=(MONTH==3))  
  
plot(TMIN.a ~ YEAR, data= station, type="p", col="grey", pch=19,  
  xlab="Year", ylab="Temperature Anomaly (C)", main="TMIN Data")  
  
TMIN.March.lm = lm(TMIN.a ~ YEAR, data= station)  
abline(coef(TMIN.March.lm), col="red" )  
  
%\item Box Plot of TMAX.
```

```
station = subset(USC00042294.anomalies[[1]])

boxplot(TMAX ~ MONTH, data= station, xlab="Month", ylab="Temperature Anomaly (C)",
        main="TMAX Data", col="grey")

)
```

3.5 KISS

Keeping it simple is critical in communicating scientific information. In this section, I try to come up with a consistent message for every state and a simple graphic.

3.5.1 Change Point Analysis

First, TMIN and TMAX and change point analysis, I need to do some background in this, but I used this method in a recent paper with co-author, but I didn't do that part of the analysis, so more reading is needed (<https://cran.r-project.org/web/packages/mcp/readme/README.html>).

3.6 Temp & Precipitation Probability

To highlight the patterns of change, it might be useful to analyze how the probability distribution might change – we can use a normal probability distribution as a theoretical distribution (and we can check if this distribution is appropriate with a Chi-Square test), or we can use the data to create an empirical distribution, which is my favored approach.

We might start with decade bins, or 20 years bins (scores) to simplify the analysis.

3.7 Using library densEstBayes

These values suggest that there is good reason

3.8 Probability Distributions

Thinking more about this one...

3.9 Evaluating Records

TBD

3.10 Export Options

TBD

4 Sea Surface Temperature Data – SURP PROJECT WAITING TO HAPPEN

In contrast to terrestrial data, sea surface temperature (SST) is quite difficult to obtain and process. There are numerous tools to access the data, but they often require knowledge of complex software tools that are not easy to set up or programming experience with python or others.

<https://climexp.knmi.nl/select.cgi?id=someone@somewhere&field=ersstv5>

There are, however, a few tools build for R users that seem to accomplish all that we need.

https://rda.ucar.edu/index.html?hash=data_user&action=register

<https://rda.ucar.edu/datasets/ds277.9/>

Alternatively, we can download flat ascII tables of gridded data:

<https://www1.ncdc.noaa.gov/pub/data/cmb/ersst/v5/ascii/>

5 Satellite Data

TBD

6 Ice-Core Data

TBD

7 Marc's Custom Functions

7.1 monthlyTrend.fun

```
## function (station)
## {
##   TMAX = station$TMAX
##   TMIN = station$TMIN
##   PRCP = station$PRCP
##   TMAX.lm = lapply(1:12, function(i) {
##     lm(TMAX.a ~ YEAR, data = subset(TMAX, MONTH == i))
##   })
##   TMIN.lm = lapply(1:12, function(i) {
##     lm(TMIN.a ~ YEAR, data = subset(TMIN, MONTH == i))
##   })
##   PRCP.lm = lapply(1:12, function(i) {
##     lm(PRCP.a ~ YEAR, data = subset(PRCP, MONTH == i))
##   })
##   TMAX.summary <- lapply(TMAX.lm, summary)
##   TMIN.summary <- lapply(TMIN.lm, summary)
```

```
## PRCP.summary <- lapply(PRCP.lm, summary)
## TMAX.stats <- lapply(TMAX.summary, function(x) {
##   c(r.squared = x$r.squared, x$coefficients[2, 1:4])
## })
## TMIN.stats <- lapply(TMIN.summary, function(x) {
##   c(r.squared = x$r.squared, x$coefficients[2, 1:4])
## })
## PRCP.stats <- lapply(PRCP.summary, function(x) {
##   c(r.squared = x$r.squared, x$coefficients[2, 1:4])
## })
## TMAX.stats <- lapply(TMAX.stats, function(x) x[c(1:5)])
## TMAX <- as.data.frame(do.call(rbind, TMAX.stats))
## TMAX$MONTH <- 1:12
## TMAX$ELEMENT <- "TMAX"
## TMIN.stats <- lapply(TMIN.stats, function(x) x[c(1:5)])
## TMIN <- as.data.frame(do.call(rbind, TMIN.stats))
## TMIN$MONTH <- 1:12
## TMIN$ELEMENT <- "TMIN"
## PRCP.stats <- lapply(PRCP.stats, function(x) x[c(1:5)])
## PRCP <- as.data.frame(do.call(rbind, PRCP.stats))
## PRCP$MONTH <- 1:12
## PRCP$ELEMENT <- "PRCP"
## return(rbind(TMAX[, c(7, 6, 2:5, 1)], TMIN[, c(7, 6, 2:5,
##   1)], PRCP[, c(7, 6, 2:5, 1)]))
## }
## <bytecode: 0xdf34e0>
```

7.1.1 Deprecated Loop Code

In case your data downloaded as Montly data, here's the code I created two years ago before the rNOAA library started failing and I had to spend all week redoing the code!

Station data frames were called GSOM (Monthly data). See if you can interpret each of the steps. Evaluate both TMAX and TMIN in GSOM by Year using MonthEvalStats() function.

```
## function (GSOM)
## {
##   sumstats = NA
##   for (m in 1:12) {
##     TMIN.lm = lm(TMIN ~ Date, GSOM[GSOM$Month == m, ])
##     TMAX.lm = lm(TMAX ~ Date, GSOM[GSOM$Month == m, ])
##     PPT.lm = lm(PPT ~ Date, GSOM[GSOM$Month == m, ])
##     sumstats = rbind(sumstats, data.frame(Month = m, Param = "TMIN",
```

```
##      Slope = coef(TMIN.lm)[2], r2 = summary(TMIN.lm)$r.squared,
##      p_value = anova(TMIN.lm)$"Pr(>F)"[1]), data.frame(Month = m,
##      Param = "TMAX", Slope = coef(TMAX.lm)[2], r2 = summary(TMAX.lm)$r.squared,
##      p_value = anova(TMAX.lm)$"Pr(>F)"[1]), data.frame(Month = m,
##      Param = "PPT", Slope = coef(PPT.lm)[2], r2 = summary(PPT.lm)$r.squared,
##      p_value = anova(PPT.lm)$"Pr(>F)"[1]))
##    }
##    sumstats = data.frame(sumstats)[-1, ]
##    rownames(sumstats) <- NULL
##    head(sumstats)
##    sumstats$Symbol = ""
##    sumstats$Symbol[sumstats$p_value < 0.05] = "*"
##    sumstats$Symbol[sumstats$p_value < 0.01] = "***"
##    sumstats$Symbol[sumstats$p_value < 0.001] = "****"
##    return(sumstats)
## }
```

8 Troubleshooting

8.1 Loading and Reading Data into the Knit Environment

Read and Load Data This function will read the csv files and load the data into R. The function will also return a list of dataframes. If you clean up the environment, you might need to read.csv data into the R environment again. Again, check the environment to see if the objects are there.

```
## function (x)
## {
##   print("This function might not be needed, waiting to see how the class does")
## }
```

Loading Saved RData files

If you have saved the RData files, you can load them into the R environment using the following code:

DOESN'T WORK YET!

```
LoadData.fun(datapath)

## [1] "RData file does not exist (Not if you have all functions in one Rmd file!)."
```

```
ls()

## [1] "datapath"          "droughtCount.fun"
## [3] "floor_decade"      "LoadData.fun"
## [5] "MonthEvalStats.fun" "monthlyTrend.fun"
## [7] "new"               "plotTrend2.fun"
## [9] "PRCP.Decade.mean"  "PRCP.Decade.sd"
## [11] "read_and_load_data.fun" "spi.fun"
## [13] "State_names"       "StateIDs"
## [15] "States"            "station"
## [17] "station_names"     "Stations"
## [19] "subtitle"          "USC00040693.anomalies"
## [21] "USC00042294.anomalies" "USC00042294.lm"
## [23] "USC00042294.post1975" "USC00042294.post1975.lm"
## [25] "USC00042294.trends"
```

The document took

9 Mapping (GIS) and Weather Stations

9.1 Simple Mapping of All US and Canada Weather Stations

```
library(here)
library(xtable)

stations.active.oldest = read.csv(
  here("04_Regional_Climate_Trends", "stations.active.oldest.csv"))

## Warning in file(file, "rt"): cannot open file '/home/mwl04747/RTricks/04_Regional_Climate_Trends/stations.active.oldest.csv': No such file or directory
## Error in file(file, "rt"): cannot open the connection
```

9.2 Map US Weather Stations

Here's map that has been transformed using a bunch of libraries that I don't really know how to use, but I found webpage that told me how to make this!

It would be nice to add Canada and Mexico so the USA is not floating in space. I'll work on that later.

```
library(usmap)
library(ggplot2)
library(sf)
```



```

## Linking to GEOS 3.7.2, GDAL 3.0.4, PROJ 6.3.2; sf_use_s2() is TRUE
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse
2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v lubridate  1.9.3      v tibble    3.2.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts()
--
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to
force all conflicts to become errors

library("rnaturalearth")
library("rnaturalearthdata")

##
## Attaching package: 'rnaturalearthdata'
##
## The following object is masked from 'package:rnaturalearth':
##
##     countries110

library(usmap)

world <- ne_countries(scale = "medium", returnclass = "sf")
class(world)

## [1] "sf"          "data.frame"

usa <- subset(world, admin == "United States of America")

(mainland <- ggplot(data = usa) +
  geom_sf(fill = "cornsilk") +
  #geom_sf(data = subset(sites), size = 3, shape = 21, fill = "gray70") +
  geom_sf(data = subset(sites.2163), size = 1, shape = 21, fill = "red") +
  coord_sf(crs = st_crs(2163), xlim = c(-2500000, 2500000), ylim = c(-2300000,
730000)))

## Error in subset(sites.2163): object 'sites.2163' not found

(alaska <- ggplot(data = usa) +
  geom_sf(fill = "cornsilk") +
  geom_sf(data = subset(sites.2163), size = 1, shape = 21, fill = "red") +
  coord_sf(crs = st_crs(3467), xlim = c(-2400000, 1600000), ylim = c(200000,
2500000), expand = FALSE, datum = NA))

```

```
## Error in subset(sites.2163): object 'sites.2163' not found

(hawaii <- ggplot(data = usa) +
  geom_sf(fill = "cornsilk") +
  geom_sf(data = subset(sites.2163), size = 1, shape = 21, fill = "red") +
  coord_sf(crs = st_crs(4135), xlim = c(-161, -154), ylim = c(18,
    23), expand = FALSE, datum = NA))

## Error in subset(sites.2163): object 'sites.2163' not found

(world.sf <- ggplot(data = world) +
  geom_sf() +
  #geom_sf(data = subset(sites), size = 3, shape = 21, fill = "gray70") +
  geom_sf(data = subset(sites.2163), size = 1, shape = 21, fill = "red") +
  coord_sf(crs = st_crs(2163), xlim = c(-2500000, 2500000), ylim = c(-2300000,
    730000)))

## Error in subset(sites.2163): object 'sites.2163' not found

mainland +
  annotation_custom(
    grob = ggplotGrob(alaska),
    xmin = -2750000,
    xmax = -2750000 + (1600000 - (-2400000))/2.5,
    ymin = -2450000,
    ymax = -2450000 + (2500000 - 200000)/2.5
  ) +
  annotation_custom(
    grob = ggplotGrob(hawaii),
    xmin = -1250000,
    xmax = -1250000 + (-154 - (-161))*120000,
    ymin = -2450000,
    ymax = -2450000 + (23 - 18)*120000
  )

## Error in eval(expr, envir, enclos): object 'mainland' not found
```

9.3 Simple Mapping of Selected State's 15 oldest, active Weather Stations

Since, we have a simple list of 15 stations in each state/territory, we can easily map them.

```
plot(my.inventory$LONGITUDE, my.inventory$LATITUDE, pch = 19, col = "red", cex = 1.5)

## Error in plot(my.inventory$LONGITUDE, my.inventory$LATITUDE, pch
= 19, : object 'my.inventory' not found
```

9.4 More Complex Mapping of Stations

```
library(geodata)
d <- worldclim_country(country = "USA", var = "tmin",
                       path = tempdir())
terra::plot(mean(d), plg = list(title = "Min. temperature (C)"))
```

10 Mapping Long-term Weather Records

10.1 Creating A Basic Trend Plot

Plot Trends by Month I have created a basic function template, but we'll develop this together when we have a chance. This function will take the list of creates a plot based on the following inputs: list of station dataframes, the element, and the month.

Function: `Function: plotTrend2.fun()`

Example of how to use the function Here's an example using the list of station dataframes anomalies. Note: I have often spelled the word anomaly wrong!

```
plotTrend.fun(USC00042294.anomalies, "TMIN", 7)

## Error in plotTrend.fun(USC00042294.anomalies, "TMIN", 7): could
not find function "plotTrend.fun"
```

Explore Results I suggest you look carefully at the plots for hidden trends. In this case, look how the temperature really starts warming after the 1960s. I don't know that that means, but seems important!

11 Extreme Events—Using Daily Records

12 Conclusions

Developing a robust method to analyze weather stations is both time consuming and difficult to justify the outcome. In part because the data suggest that each station (region) requires different types of analysis, based on the expected patterns of temperature and rainfall. As climate scientists have known for decades, the terminology of global warming is not very useful. Not because scientists are trying to hide something or promote some biased agenda, but that even

as warming of the global average is well documented, the impacts of climate change on each region is highly specific, requiring specificity in the analysis.

Hopefully, this little analysis has created some mechanism for others to appreciate this complexity.