A Comprehensive Analysis of the Continental United States' Surface Temperatures and Global Warming (1880-2017)

Daniel Foster

May 7th, 2018

Foster Scientific Consulting

6727 Bogus Avenue

San Diego, CA 92115

Contact Information:

dfoster1130@gmail.com

(619)565-4555

fosterscientificconsulting.com

Executive Summary

Abstract

Although global warming doesn't always rear its head in the continental United States in terms of temperature, many in America are still concerned about the implications of climate change here and elsewhere. Whether it be environmentally or politically, the concept of global warming is very important in the United States as Americans have a large impact on the global economy and philosophy. Global warming has only started to become a major topic of discussion within the past few decades, but trends may indicate that climate change has been occurring for decades. Additionally, spatial and temporal patterns may indicate unique variations in surface temperatures in the continental United States.

The reason behind this report is to illustrate climate change trends in the continental United States within different period of times. With this understanding based on real world data, we may be able to make powerful discoveries in the realm of global warming, actual change could start to happen to help protect our environment and the Earth.

Introduction

Description

Global warming has been a hot topic for debate for many years in both political and environmental fields. While climate change is inevitably bad for the environment and everything that lives on Earth (including humans), we are still unsure quite what to do about it. Most scientists attribute global warming to an increase in greenhouse gas emissions (mainly carbon dioxide) being released into the atmosphere as a result of industry, pollution, and a growing population. As a direct result, we have seen increased temperatures, the melting of the polar ice caps, and potential degradation of ocean habitats and more extreme weather.

In this particular report, the client is interested in a comprehensive analysis of global warming in the continental Unites States (48 states, excluding Alaska and Hawaii, abbreviated sometimes as CUS). There are already many studies that have researched and documented important data about mean surface temperatures globally, as well as centrally for different counties and in our oceans. The goal of this report is to document temperature changes over a long period of time in the Unites States, specifically between 1880 and 2017. When analyzed as a whole, we can compare specific trends during different periods of time to extrapolate a trend for future climate predictions.

Temperature anomalies are not homogenous in the United States, however, and for this reason I have chosen to use the mean surface temperature anomalies from our dataset. This should give a good basis for further understanding and scrutiny of the data. Additionally, an SVD analysis can be performed to look at both spatial and temporal effects of the data. By reviewing the data in both the space and time format, we can make assumptions about the causes of climate change in the United States and how it might relate to our oceans and other land masses. *The hypothesis is that temperature will gradually trend upwards during this period, with more major fluctuations coming within the last 20 to 30 years.*

Abstraction

The following table lists some of the important terms used in both the dataset and as a part of the analysis of the data:

Table 1: Temperature Analysis Elements			
Name	Definition	Units	Symbol
Global warming/climate change	global trending of increased temperatures and it's effects	-	-
Temperature	temperature anomalies or temperature readings/data	Celsius	℃
Latitude/Longitude	geographic global coordinates	degrees	-
Singular-value decomposition (SVD)	breaks a matrices data point into spatial and temporal characteristics	-	-
Mean(s)	a calculated average of a specific group of data	unspecified	*m/M*
Spatial/Temporal	data in space and time format	-	*U and V*

A temperature analysis should be easy to understand, as it effects everybody and is important for the environment and how we live our daily lives. An effort will be made to make the conclusions and results of this report accessible to scholars of all types, while making it robust and comprehensive enough so that scientists can make their own conclusions from the presented findings.

Data and Method

Equations for a Mathematical Model

The dataset that will be used for the report is the NOAA Global Temperature dataset with global land and ocean monthly gridded (5 degree by 5 degree, latitude/longitude) temperature anomalies from 1880 to 2017, all relative to a 1971-2000 climatology. The dataset is in .asc (*ActionScript*) format and is unruly as it contains a very large amount of raw data, much of which is "bad" data because of incorrect measurements (i.e. null values). The strength of the dataset is that it includes data for the entire globe for small variations in both space (latitude/longitude, 5 degree gridded) and time (by month). None of the "bad" data was used in any calculations.

The data analysis for the report is done using R Studio and the R source code is included at the end of the report so that anybody can reproduce the figures or modify the calculations. The following methods were used for analysis of the data:

1. Prepare the data for analysis and constrain to the correct time period and spatial region.
2. Filter out missing (null) values.
3. Calculate yearly mean temps for the continental United States from 1880-2017 (plots will be done with titles and annotations using the line graphs function, "plot").
4. Analyze the slope of the line to determine a trend and determine a trend line for the last few decades, using the linear model function ("lm").
5. Complete an SVD analysis; analyze eigenvectors/eigenvalues and spatial and temporal data, using the built in SVD tool in R (i.e. test_data=svd(data), U=test_data$u).
6. Make conclusions about climate change in the continental Unites States.

Solutions

A linear regression model will be used as a best-fit trend line for the mean surface temperature anomalies from 1880 to 2017. The model used for the SVD portion of the analysis will be the basic SVD matrix space-time model with space-time components, represented by U and V, respectively. And, while R is a powerful tool capable to doing large-scale data calculations and manipulations, the following two equations/models were used directly to determine a solution to the problem, although many other equations have an influence on climate change (such as carbon emissions or natural-gas depletion):

- SVD (linear algebra) – A = UDV': used in a space-time decomposition for vectors U and V, in space and time format, respectively.
- Climate Model (not used in calculations, but useful as a baseline model) – $T = \sqrt[4]{\frac{(1-a)S}{4\epsilon\sigma}}$: where a (Earth's albedo), S(solar constant), ϵ (emissivity), and σ (the Stefan-Boltzmann constant) help determine T, the relative temperature of the Earth's surface/atmosphere at a given time.

The following data table gives an example of the data after steps 1 and 2 above (where the data now represents monthly mean surface temperatures for the given latitude/longitude range), which can be used in linear regression and to find trends:

Table 2: Sample Data – Mean Temperature Anomalies (continental United States)

Month	Temperature Anomaly (℃)
1	-0.01976
2	-1.42324
3	0.147177
4	-0.51713
5	-0.45798

In the table above, months represent a month in the range of 1 to 1644 ((2017 – 1880) * 12 months) and temperature anomalies represent monthly surface temperature means in the continental United States as compared to the standard comparative climatology. *Note:* If a 0 value is displayed in the temperature anomaly column, this value is seen as a null or misrepresented value.

Results

Solutions & Interpretation of Modeling Results

The logical first step when analyzing the dataset on any domain or location is to view the data model over time. It was at first difficult to prepare the data to be shown in this manner, but by manipulating the latitude and longitude separately, I chose the continental United States as the area of interest. The figure below (Figure 1), shows the time versus mean temperature anomaly plot for the continental United States with approximate latitudes between 30 and 50 degrees and longitudes between 230 and 295 degrees (all relative to the NOAA scale). Also of note in the figure (and all data in this report), is that all temperatures are in degrees Celsius. The temperature anomalies in this figure were found by taking the mean of all temperature anomalies in a given month for the correct latitudinal/longitudinal range. The figure examines three time periods of interest where global warming and climate change can be evaluated based on trends and data.

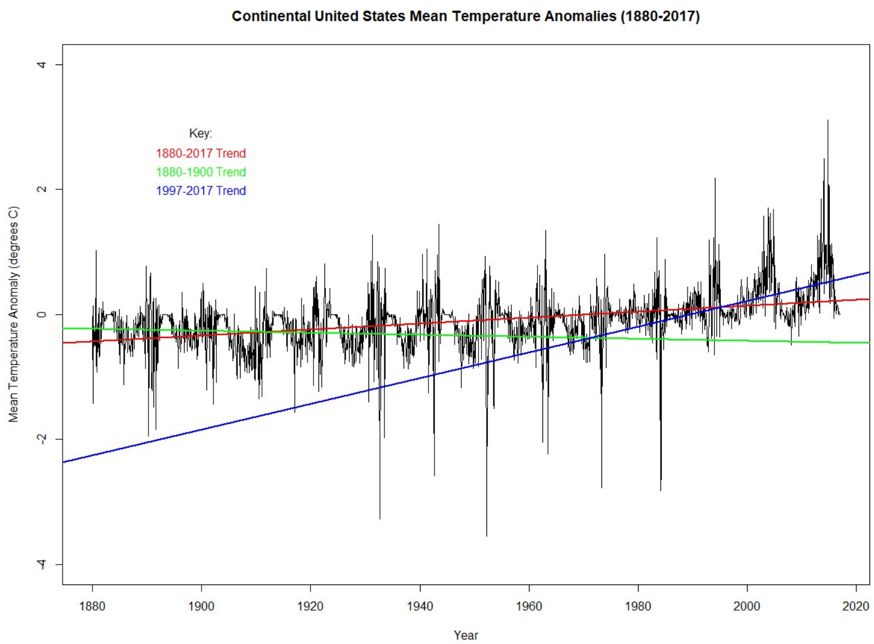

Figure 1: Continental United States Temperature Anomalies and Trends

By first looking at the figure, you can see an obvious shift or trend upwards towards the start of the 21st century. Without any other knowledge about the data whatsoever, this by itself tells a lot about why climate change has been such a hot topic. The trend line (found by linear regression) for the time period from 1880-2017 has a slope of 0.004777. This means that for every data point on the graph, the temperature anomaly associated with it increases, on average, by 0.004777 degrees Celsius. This may not seem like much, but when calculating the total increase in temperature over the 137 years, this adds up.

Next, I analyzed the two most opposite periods of the model to examine their trends. I chose two 20 year periods; one period is from 1880-1900 and the other is from 1997-2017. As an example, I thought by examining these two periods, I could contrast the difference in their trend lines. The 1880-1900 timeframe trend line is highlighted in green and shows a negative slope of -0.001615. This confirms that during this period, temperature anomalies were trending downwards relative to the climatology from NOAA. In contrast, the 1997-2017 period, displayed in blue, illustrates the greatest temperature anomaly change in the positive quadrant. During this period, the trend line suggests a slope of 0.02059. As you can see, the blue line shows a very large migration from either the total data trend or the 1880-1900 trend, suggesting the fact that climate change has in fact occurred starting only recently. The only data not pictured here is data from the future, which could trend down again to meet up with the 137 year average, or continue the upward trend which many scientists and climatologists suggest, which would certainly indicate the global warming is in fact real.

In regards to this data, a sensitivity analysis is included below, in Table 3.

Table 3: Sensitivity Analysis – Time vs. CUS Temperature Anomaly Trend

Time Period	Trend Slope (℃)
1880-2017 (whole dataset)	+0.004777
1880-1900 (first 20 years)	-0.001615
1997-2017 (last 20 years)	+0.02059
1880-1945 (first half of data)	+0.001178
1946-2017 (second half of data)	+0.01152

The sensitivity analysis shows that the data is pretty sensitive to any change in the time period. Essentially, the trends are increasing as the time increases, showing that while the data in sensitive to the independent variable, it is only sensitive to the independent variable and that mean surface temperature anomalies are increasing rapidly. It could be insinuated that for other regions of the globe, such as the polar regions, this small increase in temperature could have a very large impact on the environment.

Now, with the data analysis shown previously, it is smart to turn our attention to a space-time analysis of the data using SVD. By viewing the spatial and temporal components of the datum separately, hopefully global warming trend can be more conclusive.

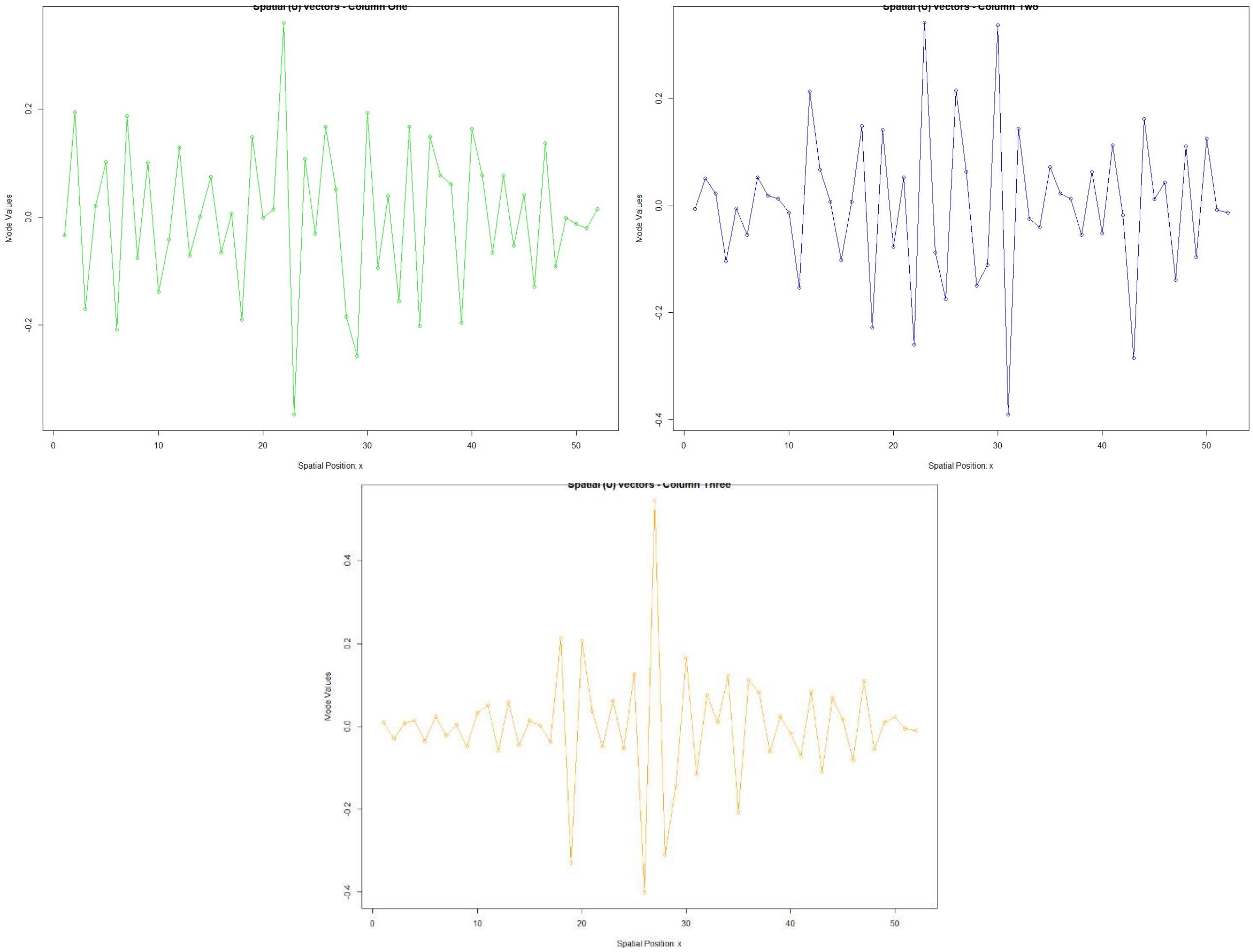

Figure 2: Spatial Vectors (first three columns) for the Continental United States EOF's

The spatial decomposition of the data, using the first three vector columns, all show peak(s) near the middle of the plot, suggesting the spatial relevance of the data is relatively consistent, and, at least at this point, we can't determine if the temporal analysis might show more insight. These peaks in the EOF graphs may indicate events based on spatial patterns, but don't seem to indicate any obvious correlation to the ideas of climate change or global warming in the United States.

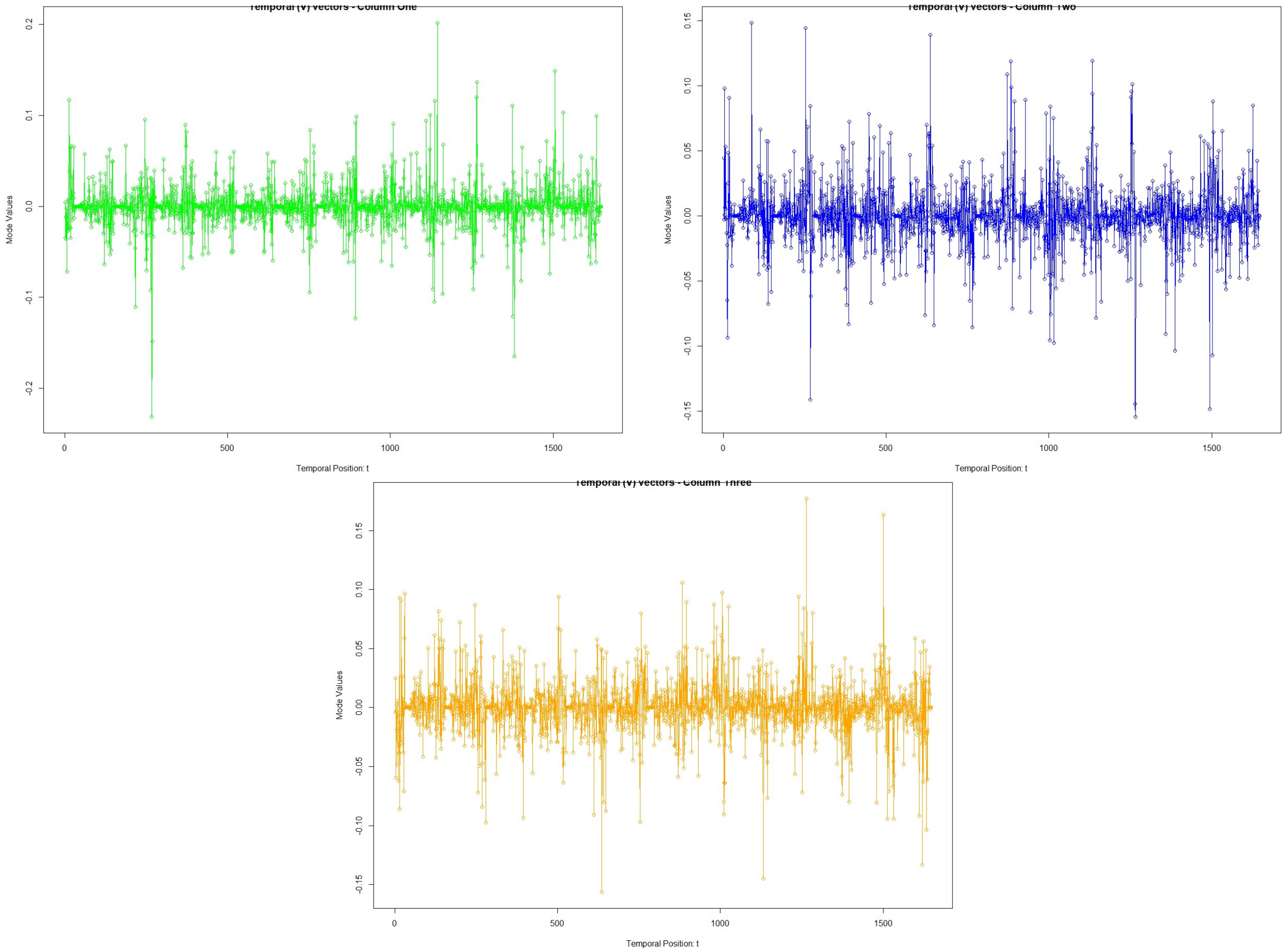

Figure 3: Temporal Vectors (first three columns) for the Continental United States PC's

The temporal vectors do a better job of explaining the dataset in terms of time. The peaks suggest changes in temperature anomalies based on time, which, as discussed earlier, was our hypothesis that time has the biggest effect on climate change. These plots don't give insight as to the reasoning behind global warming, but do in fact give proof that, at least in the continental United States, climate change is definitely taking place and could very well be attributed to an increase in current environmental issues such as greenhouse gas emissions and excessive pollution based on population increases.

Conclusion and Discussion

After creating a model and utilizing the powerful tool that is R to interpret the required data, it is obvious that climate change is taking place in the continental United States. A comprehensive analysis was done to determine the differences in mean surface temperature anomalies based on time fluctuations over the past 137 years, and a sensitivity analysis gave additional insight into the fact that the last 20 year period where temperature anomaly fluctuations were negative were before the year 1900. The sensitivity analysis proves that while the data is extremely sensitive to change in time, something is influencing the effect that time has on temperature, which would seem to insinuate the common causes of global warming reasoning.

An SVD analysis was also performed to interpret the data in a new domain. Spatial (U) and temporal (V) vectors were plotted to see if any obvious anomalies could be found. Again, as was true with the first calculations, the temporal part of the datum is much more relevant than the spatial. One reason the spatial data also may not be relevant is that only the continental United States was examined, and because the CUS is very much similar in its climatology (at least compared to some other countries, such as China or Russia), spatial patterns may not be obvious, or at least very similar.

In conclusion, I believe that climate change and global warming is in fact very real and will most likely follow the 1997-2017 trend line. An increase in only quarter of a degree in temperature per year would ultimately result in an increase in 25 degrees over the next century. Although this seems crazy, we very well may be hurtling towards this horrible fate. Hopefully, environmental steps can be taken to reduce the effects of this global warming phenomenon and ultimately resolve the issue that is horrible pollution and misuse of the land, water, and air that was given to us. If we don't do something to change the increasing trend, we may not have a habitable Earth to pass on to the next generation.

I would very much like to continue research in this subject. What is the chemical reasoning behind this global warming trend? How does it affect biological ecosystems and the lives of animals? Why are trends so different for areas of land such as the continental United States when compared to water areas such as the Pacific Ocean? If more research and information can be presented on these important topics, we can start to battle global warming and rising temperatures.

References

"Global Climate Change." *NASA*, NASA, 3 June 2014, climate.nasa.gov/. (*current conditions via NASA*)

"Global Temperature Trends From 2500 B.C. To 2040 A.D." *Global Temperature Trends Since 2500 B.C.*, www.longrangeweather.com/global_temperatures.htm. (*climate change trends*)

Mann, Michael E., and Henrik Selin. "Global Warming." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 13 Apr. 2018, www.britannica.com/science/global-warming. (*global warming general information*)

Mills, Peter. "Singular Value Decomposition (SVD) Tutorial: Applications, Examples, Exercises." *Stats and Bots*, Stats and Bots, 5 Oct. 2017, blog.statsbot.co/singular-value-decomposition-tutorial-52c695315254. (*SVD information and application*)

"NOAA Merged Land Ocean Global Surface Temperature Analysis (NOAAGlobalTemp)." *National Climatic Data Center*, www.ncdc.noaa.gov/data-access/marineocean-data/noaa-global-surface-temperature-noaaglobaltemp. (*dataset citation*)

R Source Code

```
# Final project R source code for all data and figures.

# NOAA Global Mean Surface Temperature Anomalies (1880-2017)

rm(list = ls(all = TRUE))

# Read in .asc file (unopenable).

noaa_initial_data = scan("C:/Users/Dan/Desktop/336
Documents/NOAAGlobalTemp.gridded.v4.0.1.201802.asc")

length(noaa_initial_data) # Initial read length check.


# Hard to maniulate (x vs. y?).

noaa_initial_data[1:3]

#month, year, temp (72 rows, 36 columns)

data_samp1 = seq(1, 4267129, by = 2594)

data_samp2 = seq(2, 4267130, by = 2594)

length(data_samp1)

length(data_samp2)


month1 = noaa_initial_data[data_samp1] # Sample (months).

year1 = noaa_initial_data[data_samp2] # Sample (years).

head(month1)

head(year1)

length(month1)

length(year1) # Dimension check  based on year alone.


var_names <- paste(year1, sep = "-", month1) # year-month

head(data_samp1)

head(data_samp2)

data_samp3 = cbind(data_samp1, data_samp2) # Column bind.

data_samp4 = as.vector(t(data_samp3))

head(data_samp4)
```

```
data_next <- noaa_initial_data[-data_samp4] # Remote months/years.
length(data_next) / (36 * 72) # 137 years (2017-1880), (137 * 12) = 1,645 months

data_last <- matrix(data_next, ncol = 1645)
colnames(data_last) <- var_names # Bind column names.
latitudinal_data = seq(-87.5, 87.5, length = 36)
longitudinal_data = seq(2.5, 357.5,  length = 72)
LATITUDE = rep(latitudinal_data, 72)
LONGITUDE = rep(longitudinal_data[1],36)

for (i in 2:72) {LONGITUDE = c(LONGITUDE, rep(longitudinal_data[i], 36))}
noaa_data = cbind(LATITUDE, LONGITUDE, data_last)
head(noaa_data) # Gives temps based on lat/lon.
dim(noaa_data) # NOAA final dim sanity check.

for(x in 1:nrow(noaa_data)){
  for(y in 3:ncol(noaa_data)){
    if(noaa_data[x, y] < -300)
      noaa_data[x, y] <- 0
  }
}

# Now manipulate data for final intgeration with correct columns.
# Time - Year
time_year = seq(1880, 2017, by = 1) # 1880-2017
time_month = rep(time_year, each = 12)
time_by_month = rep(1:12, 138)
data_samp5 = paste(time_year, "-", time_by_month)
data_samp6 = c("Latitude", "Longitude", data_samp5)
length(data_samp6)
```

```
colnames(noaa_data) <- data_samp6[1:1647] # Transpose column names and remove 2017 last (extra) months.

# Plot field data via maps lib.
library(maps)
lat2_vec = seq(-87.5, 87.5, length = 36)
lon2_vec = seq(2.5, 357.5, length = 72)

vectors_to_map = noaa_data[,1635]
vector1 = pmin(vectors_to_map, 6)
vector2 = pmax(vector1, -6)
mapmat_vectors = matrix(vector2, nrow = 72)
plot(seq(-90, 90, len = 36), mapmat_vectors[36, c(1:36)], type = "l")

# Select latitude and longitude (here 30 to 50 and 230 to 295, respectively). CUS.
lat_and_lon <- which(noaa_data[,1] > 30 & noaa_data[,1] < 50 & noaa_data[,2] > 230 & noaa_data[,2] < 295)
dim(lat_and_lon)
noaa_manip = matrix(0, nrow = 52, ncol = ncol(noaa_data)) # nrow = size of lat_and_lon

# Final cleaned data for all NOAA stuff.
for(a in 1:52){
  noaa_manip[a,] = noaa_data[lat_and_lon[a],]
}

dim(noaa_manip)
#write.csv(noaa_manip, "C:/Users/Dan/Desktop/data.csv") To view data.
#mtf_dates = seq(1880, 2017.1666666666, by=0.0833333333) Don't need to include extra data in 2017.
#plot(mtf_dates)
```

```
svd_data = svd(noaa_manip)

dim(noaa_manip)


U = svd_data$u

D = svd_data$d

V = svd_data$v


### Annual SAT of EOF's and PC's (variance).

data_temp=seq(855,1466,by=12)

annual_SAT=matrix(0,nrow=52,ncol=50) # 52 rows.

for (i in 1:50) {annual_SAT[,i]=rowMeans(noaa_manip[,seq(data_temp[i],data_temp[i+1]-1,
length= 12)])}

annual_SAT2=svd(annual_SAT)


annual_data <- annual_SAT2$v[,1]

dim(annual_data)

time_seq = seq(1951, 2000)

plot(time_seq,-annual_data,type='o')


### Mean temperature analysis.

par(mfrow=c(1,1))

USland = colMeans(noaa_manip)

write.csv(USland, "C:/Users/Dan/Desktop/data.csv")

write.csv(mtf_dates, "C:/Users/Dan/Desktop/mtf_dates.csv")

mtf_dates=seq(1880,2017,by=0.0833333333)

plot(mtf_dates,USland[3:1647],type='l',xlab="Year",ylab="Mean Temperature Anomaly
(degrees C)",xlim=range(1880:2017),ylim=range(-4:4),main="Continental United States Mean
Temperature Anomalies (1880-2017)")

lm1=lm(USland[3:1647]~mtf_dates,na.action=na.omit) # Full regression.

lm1

abline(lm1,col='red',lwd=2)
```

```
text(1900,2.9,col='black','Key:',cex=1) # Key.
text(1900,2.6,col='red','1880-2017 Trend',cex=1)


lm2=lm(USland[3:242]~mtf_dates[1:240],na.action=na.omit) # 1880-1900
lm2
abline(lm2,col='green',lwd=2)
text(1900,2.3,col='green','1880-1900 Trend',cex=1)

lm3=lm(USland[1407:1647]~mtf_dates[1405:1645],na.action=na.omit) # 1997-2017
lm3
abline(lm3,col='blue',lwd=2)
text(1900,2,col='blue','1997-2017 Trend',cex=1)

lm4=lm(USland[3:818]~mtf_dates[1:816],na.action=na.omit) # 1880-1945
lm4

lm5=lm(USland[833:1647]~mtf_dates[831:1645],na.action=na.omit) # 1946-2017
lm5

# Print the first ten eigenvalues of the space-time decomposition.
D[1:10]
# [1] 1924.77727  114.59196  101.85219   94.19324   67.52522   55.64969   52.48308   41.18995
40.87237
#[10]   36.59311

plot.new()
par(mar = c(4, 4, 0.2, 0.5))

# Spatial data (in terms of x).
```

```
dim(U)

plot(1:52, U[,50], type = "o", col = "green", xlab = "Spatial Position: x", ylab = "Mode Values",
ylim = c(min(U[,50]), max(U[,50])), lwd = 1.5, main = "Spatial (U) vectors - Column One")

plot(1:52, U[,51], type = "o", col = "blue", xlab = "Spatial Position: x", ylab = "Mode Values",
ylim = c(min(U[,51]), max(U[,51])), lwd = 1.5, main = "Spatial (U) vectors - Column Two")

plot(1:52, U[,52], type = "o", col = "orange", xlab = "Spatial Position: x", ylab = "Mode Values",
ylim = c(min(U[,52]), max(U[,52])), lwd = 1.5, main = "Spatial (U) vectors - Column Three")


# Temporal data (in terms of t).

dim(V)

plot(1:1647, V[,52], type = "o", col = "green", xlab = "Temporal Position: t", ylab = "Mode
Values", lwd = 1.5, main = "Temporal (V) vectors - Column One")

plot(1:1647, V[,51], type = "o", col = "blue", xlab = "Temporal Position: t", ylab = "Mode
Values", lwd = 1.5, main = "Temporal (V) vectors - Column Two")

plot(1:1647, V[,50], type = "o", col = "orange", xlab = "Temporal Position: t", ylab = "Mode
Values", lwd = 1.5, main = "Temporal (V) vectors - Column Three")
```