Final Project

Abbas Paryavi, Tyler Fitzwater and Jack Wesolowski

An Analysis of Global Warming

While the climate has shifted thorughout history, between eras of warmer and colder climates, in the past century, scientific organizations have begun to ring the alarm bell on global warming. The climate has a direct effect on our quality of life. As water levels change, weather grows harher, and food becomes harder to grow, everyone will feel the impact of a changing climate.

We have decided to base our final project on issues of global warming and its effects on the environment. We will highlight factors that have changed such as percipitation patterns around the world. We have gathered data from the United States Environmental Protection Agency and NOAA to analyse the patterns of global warming and precipitation throughout the decades. However, since we were not able to find appropriate global precipitation data, we limited our precipitation analysis to the United States and all of the states inside the US.

We believe that we are going to be see subtle changes in some parts of the country but some more drastic changes in precipitation in other regions. This is because the global warming has caused more dramatic temperatures by giving us colder winters and hotter summers. However, places were temparature is fairly constant throughout the year and precipitation was never that much, we believe might not be effected by global warming as much. Now let's see if our hypothesis is correct.

Data Gathering and Tidying

Before we get into it, all of the libraries needed will need to be imported and will do so at the very beginning so people can see what libraries are being used. The most important ones being pandas, matplotlib, and for our writeup JSON. The rest are mostly just allow quality of life and efficiency improvements, so that the reader can follow the idea.

For tidying our data, we check for NaN and to make sure that our data is ready for analysis. For example removing unnecessary columns or removing outliers.

In [3]:
```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib._color_data as mcd
import numpy as np
import json
import requests
from json import JSONEncoder
from matplotlib.pyplot import figure
```

Climate Change

Then we will use the global climate data obtained from NOAA, which are also available as downloadable csv files.
Links to both the downloaded files, and the source will be included in the references section. First we will need to load the data from the comma separated values files from NOAA, and then put them into a pandas database. We can figure out what to do for data wrangling and such after this step from the resulting dataframe. For this step, we will gather all of the temperature data for the global, and the regions of the globe over the years. All of this data is easily reachable on the NOAA source using these same parameters:

Region: (each region)

Timescale: 1-Month

Month: January

Start Year: 1880

End Year: 2022

Surface: (only one option available for each region)

Here is the code for this step:

In [5]:

```
# this function is specifically to parse the NOAA data!
def json_file_to_df(path):
  # read data from file
  df_file = open(path)
  obj = json.load(df_file)
  df_file.close()
  data = obj["data"]
  region = obj["description"]["title"].split(" Temperature")[0] # get region tit
  df = pd.DataFrame.from_dict(data, orient="index")
  # label columns
  df.reset_index(inplace=True)
  df.rename(columns={"index":"year", 0:"jan_temp_celsius"}, inplace=True)
  df["year"] = pd.to_numeric(df["year"])
  df["jan_temp_celsius"] = pd.to_numeric(df["jan_temp_celsius"])
  return df, region


# Read data from global file
global_temp_df, region = json_file_to_df("./content/global.json")
# global_temp_df.style.set_caption("Global")
print(region)
global_temp_df
```

```
Global Land and Ocean
```

Out[5]:

	year	jan_temp_celsius
0	1880	-0.05
1	1881	-0.08
2	1882	0.09
3	1883	-0.27
4	1884	-0.22
...	...	...
138	2018	0.75
139	2019	0.93
140	2020	1.13
141	2021	0.78
142	2022	0.88

143 rows × 2 columns

Additionally, we will get the regional climate data in the same way. Having the regional data will show that there isn't some outlier that is causing a skew in the global averages. If we show that the pattern is relatively the same as the global pattern for each/all regions, then we can further show that the earth as a whole is warming and not just some region(s).

In [7]:

```
filenames = ["africa.json", "asia.json", "atlantic.json", "caribbean.json", "eur
colors = ["xkcd:purple", "xkcd:green", "xkcd:blue", "xkcd:pink", "xkcd:brown", "
regions = []
climate_dfs = []

# read data from file names in filenames array
for name in filenames:
  file_path = f"./content/{name}"
  df, r = json_file_to_df(file_path)
  regions.append(r)
```

```
  climate_dfs.append(df)

# display regions captured
print(f"{regions}\n")

# display first df (africa)
print(regions[0])
climate_dfs[0].head()
```

```
['Africa', 'Asia', 'Atlantic MDR', 'Caribbean Islands', 'Europe', 'North Americ
a', 'Oceania', 'East N Pacific', 'South America']

Africa
```

Out[7]:

	year	jan_temp_celsius
0	1910	-0.59
1	1911	-0.62
2	1912	-0.29
3	1913	0.05
4	1914	0.74

Here we see that all regions were captured.

In [8]:

```
# display last df (south america)
print(regions[-1])
climate_dfs[-1].head()
```

```
South America
```

Out[8]:

	year	jan_temp_celsius
0	1910	-0.44
1	1911	-0.48
2	1912	-0.17
3	1913	-0.35
4	1914	-0.17

Precipitation

We will gather the precipitation data from the United States Environmental Protection Agency. The dataset can be downloaded as a comma separated values (.csv) file, and will be included here. These files can be very easily parsed with pandas using a one liner.

In [10]:

```
# read precipitation data from csv file
precipitation_df = pd.read_csv("./content/sample_data/pr_timeseries_annual_cru_1
precipitation_df.head()
```

Out[10]:

	Unnamed: 0	United States	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	De
0	1901	632.77	1384.46	438.15	214.09	876.11	557.73	361.58	1424.90	
1	1902	694.42	1282.48	439.00	204.91	1319.22	614.28	365.83	1322.01	
2	1903	661.39	1272.60	415.69	197.38	1096.21	503.50	354.56	1303.55	
3	1904	627.57	1010.58	402.31	216.57	1041.74	701.05	377.94	1078.13	
4	1905	733.08	1443.03	430.50	489.95	1498.87	529.42	456.82	958.67	

5 rows × 53 columns

In [11]:

```
# change first column name to appropriate Year name
precipitation_df.rename(columns = {'Unnamed: 0':'Year'}, inplace = True)
precipitation_df.head()
```

Out[11]:

	Year	United States	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delawa
0	1901	632.77	1384.46	438.15	214.09	876.11	557.73	361.58	1424.90	1145.6
1	1902	694.42	1282.48	439.00	204.91	1319.22	614.28	365.83	1322.01	1281.7
2	1903	661.39	1272.60	415.69	197.38	1096.21	503.50	354.56	1303.55	1200.1
3	1904	627.57	1010.58	402.31	216.57	1041.74	701.05	377.94	1078.13	991.0
4	1905	733.08	1443.03	430.50	489.95	1498.87	529.42	456.82	958.67	1077.3

5 rows × 53 columns

We checked for NaN and all of our data had every single cell full, therefore we didn't need to change anything on that part. We changed some header names and columns to make our data best fit for what we want to do. However, other than these we didn't really have to do any thing else, becasue the datasets were fairly ready itself.

Data Analysis

Climate Map

To be able to see if there is a correlation between precipitation amount and temperature, first we need to see if there is an overall increase in the climate patterns. To do this we will graph the data for temperature over year for global, and for regions of the globe. Having the regional data on top of the global data is very important because it shows there isn't some outlier that could have caused a skew in the averages increasing over year. As long as the regional data has a similar distribution pattern, aside from tightness or variance, and min/max values, then we can show that every region is contributing to the increases and changes to the global climate.
To learn more about scatter plots, click Scatter Plots, and to learn more about regression lines click Regression Lines.

In [12]:

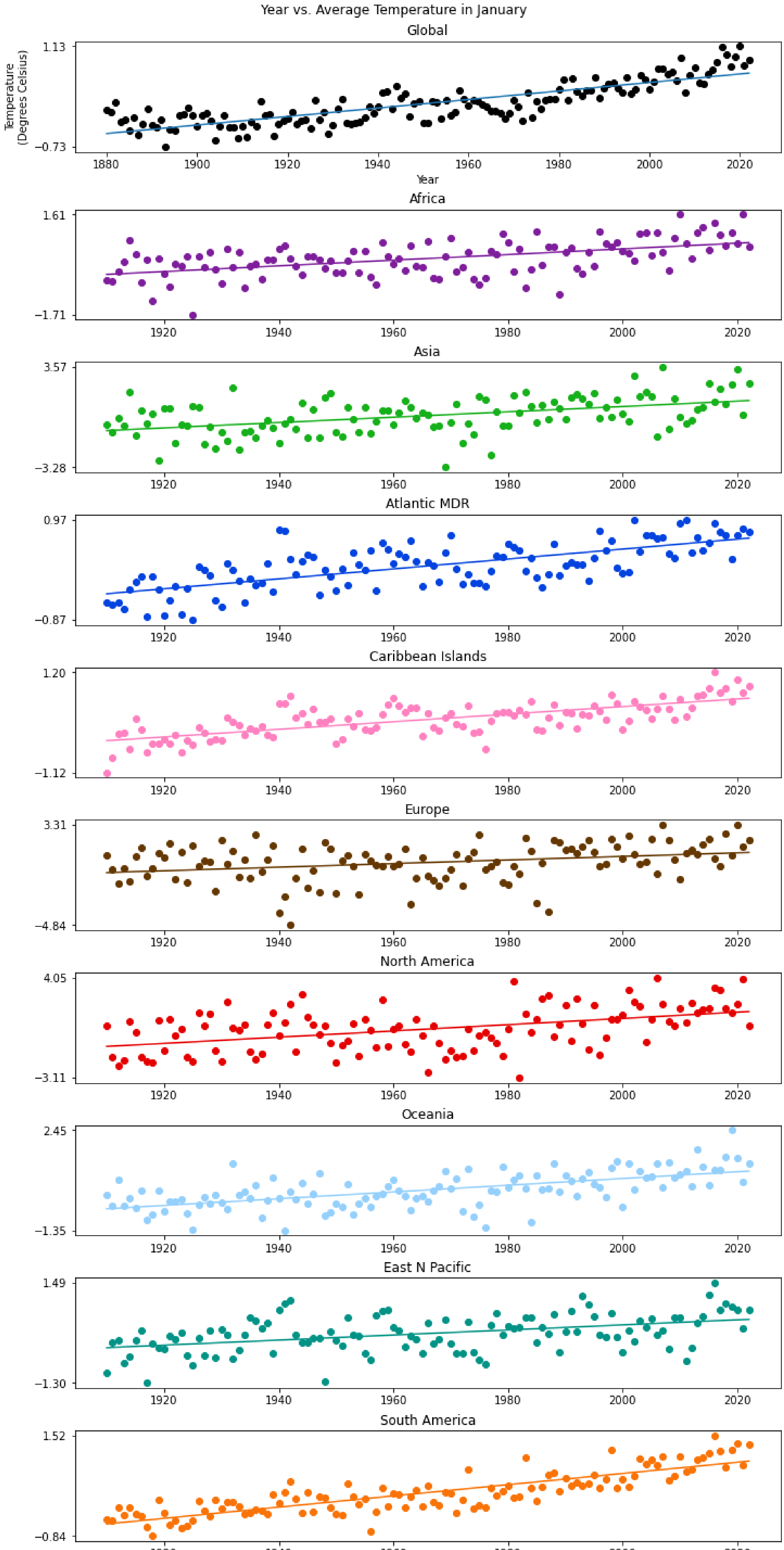
```
# colors
color_map = mcd.XKCD_COLORS # dict for: color names -> hex code
```

In [13]:

```
fig, ax = plt.subplots(len(climate_dfs) + 1, constrained_layout=True)
fig.suptitle("Year vs. Average Temperature in January")
fig.patch.set_facecolor("white")
fig.set_size_inches(10, 20)
# fig.tight_layout()
ax[0].set_xlabel("Year")
ax[0].set_ylabel("Temperature\n(Degrees Celsius)")

# global
min_temp = global_temp_df["jan_temp_celsius"].min(axis=0)
max_temp = global_temp_df["jan_temp_celsius"].max(axis=0)
yticks = [min_temp, max_temp]
x = global_temp_df["year"]
y = global_temp_df["jan_temp_celsius"]
# get slope and y intercept of the temp vs year
coef = np.polyfit(x, y, 1)
poly1d_fn = np.poly1d(coef)
# do a scatter plot with x being the year and y beging the temp
ax[0].scatter(x, y, color=color_map["xkcd:black"], label="Global")
ax[0].plot(x, poly1d_fn(x))
ax[0].set_yticks(yticks)
ax[0].set_title("Global")

# regions
for r in range(len(climate_dfs)):
  # do the same thing as for the global plot except we are doing it
  # in a loop for all regions
```

```
    min_temp = climate_dfs[r]["jan_temp_celsius"].min(axis=0)
    max_temp = climate_dfs[r]["jan_temp_celsius"].max(axis=0)
    yticks = [min_temp, max_temp]
    x = climate_dfs[r]["year"]
    y = climate_dfs[r]["jan_temp_celsius"]
    color = color_map[colors[r]]
    coef = np.polyfit(x, y, 1)
    # print slope of temperature increase for each region
    print(regions[r], ' annual temp increase:\n', coef[0])
    poly1d_fn = np.poly1d(coef)
    ax[r + 1].scatter(x, y, color=color, label=regions[r])
    ax[r + 1].plot(x, poly1d_fn(x), color)
    ax[r + 1].set_yticks(yticks)
    ax[r + 1].set_title(regions[r])
```

```
Africa  annual temp increase:
 0.009442411338079722
Asia  annual temp increase:
 0.018268181515736268
Atlantic MDR  annual temp increase:
 0.009103732783285651
Caribbean Islands  annual temp increase:
 0.008773704171934248
Europe  annual temp increase:
 0.014601104531239646
North America  annual temp increase:
 0.022330577550069914
Oceania  annual temp increase:
 0.01274926808170871
East N Pacific  annual temp increase:
 0.007043382793266341
South America  annual temp increase:
 0.013188502229023891
```


Year vs. Average Temperature in January
Global
Temperature
(Degrees Celsius)
1.13
−0.73
1880
1900
1920
1940
1960
1980
2000
2020
Year
Africa
1.61
−1.71
Asia
3.57
−3.28
Atlantic MDR
0.97
−0.87
Caribbean Islands
1.20
−1.12
Europe
3.31
−4.84
North America
4.05
−3.11
Oceania
2.45
−1.35
East N Pacific
1.49
−1.30
South America
1.52
−0.84

Regional Temperature Increase Analysis

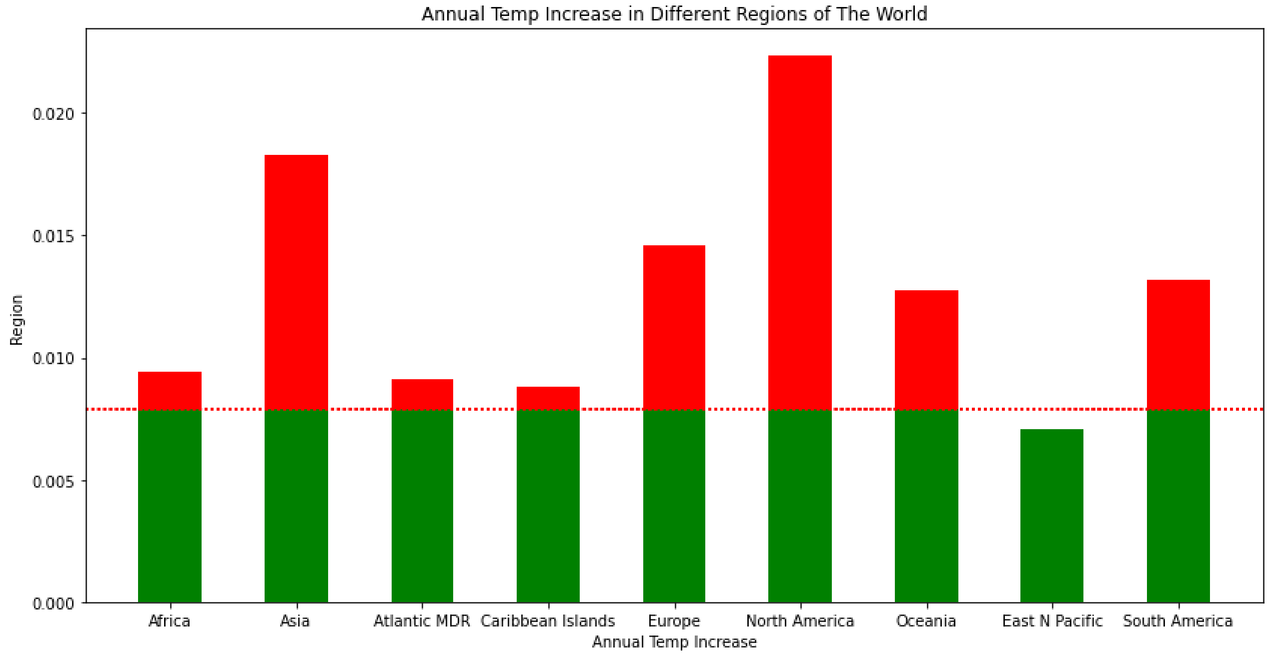
These graphs give us a very good insight on what is happening with the temperature around the world. It is obvious that some regions of the world have a more dramatic increase in temperature as opposed to some others. To better understand the temperature increase difference, we can also see the exact average increase in temperature per year in eac region before the graphs. For example, North America has the absolute highest increase in temperature and Asia is next. However, places like the Caribbean Islands or Atlantic MDR have some of the least increasing temperatures around the world (but they're still increasing). We can definetly see a correlation in the region since they both are very close if not on the equator line. And North America is over the equator line that is why it has the most extreme effect on there. Based on the article *Which parts of the planet are warming the fastest, and why?*, "The earth's largest land masses and its north and south poles are warming the fastest, mainly because of differences in how these areas reflect energy from the sun" (Climate Portal).

Bar Plot of Regional Temperature Increase

We are also going to take a different approach for looking at the temperature rise to get a better understanding of the regions most effected. We are going to make a barplot that shows all of the regions' average annual temperature increase and it's difference with the global average. We could have used soome other sort of plot to show this relationship such as heatmap, but bar plots show a much better insight for what we want to achieve in this section which is to see the difference in temperatrue change in different regions. To read more on barplots click Bar Plot.

In [14]:

```
# set the size of the plot to 14 width and 7 height
fig = plt.figure(figsize = (14, 7))

# get the slope of global temperature increase
x = global_temp_df["year"]
y = global_temp_df["jan_temp_celsius"]
globalSlope, b = np.polyfit(x, y, 1)

# regions
for r in range(len(climate_dfs)):
  # get average temperature increase in each region and set it to m
  min_temp = climate_dfs[r]["jan_temp_celsius"].min(axis=0)
  max_temp = climate_dfs[r]["jan_temp_celsius"].max(axis=0)
  yticks = [min_temp, max_temp]
  x = climate_dfs[r]["year"]
  y = climate_dfs[r]["jan_temp_celsius"]
  color = color_map[colors[r]]
  m, b = np.polyfit(x, y, 1)

  # upper half of the regional bar. Remainder of temp increase passed the global
  athreshold = m - globalSlope if m - globalSlope > 0 else 0
  # lower half of the regional bar. Temp increase of each region
  bthreshold = m
  # creating the bar plot
  # lower half colored green
  plt.bar(regions[r], bthreshold, color ='green',
        width = 0.5)
  # upper half colored red
  plt.bar(regions[r], athreshold, color ='red',
        width = 0.5,bottom=globalSlope)

  plt.xlabel("Annual Temp Increase")
  plt.ylabel("Region")
  plt.title("Annual Temp Increase in Different Regions of The World")

  # put dotted line indicating global average
  plt.axhline(globalSlope, color='red', ls='dotted')
```



Regional Temperature Increase Analysis with Barplot

Looking at the barplot, we can easily see the differences in temperature change in different regions of the world. The red dotted line is the global average. As we can analysed in the previous section North America and Asia have the two highest increase in temperature in the entire wolrd. This shows how dramatically temperatures are changing in some parts of the world.

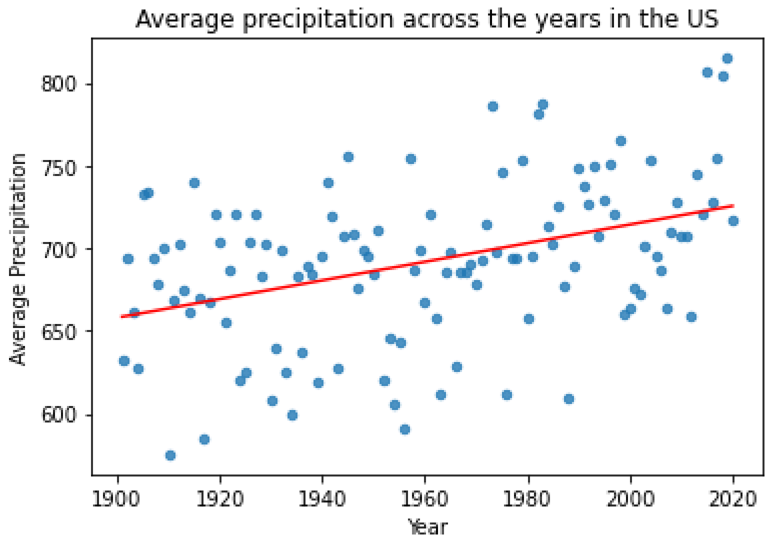
Top two:
North America - 0.02233
Asia - 0.01826
Bottom two:
East N Pacific - 0.00704
Caribbean Islands - 0.00877

Next we are going to look at what effects this temperature change has on the precipitation in the US. We are going to be looking at different parts of the country.

Precipitation Map

Before we graph the precipitation vs temperature graph, we first need to do some analysis of our precipitation data. We are going to graph the US precipitation change from around 1900-2020. This way we can see any approximate changes in precipitation over the years. This can give us some insight on what we can expect from our final analysis of precipitation vs temperature.

In [15]:
```
# plotting a scatter graph of US precipitation vs Year
precipitation_df.plot(x="Year", y="United States", kind="scatter",alpha=0.8)
plt.ylabel('Average Precipitation')
plt.title('Average precipitation across the years in the US')

x = precipitation_df['Year']
y = precipitation_df['United States']
#obtain m (slope) and b(intercept) of linear regression lineir
m, b = np.polyfit(x, y, 1)

#add linear regression line to scatterplot
plt.plot(x, m*x+b, color='red')
plt.show()
```



Here we can see some change in precipitatoin as the years go by. From the scatter plot we see some very random points in the graph, however, there is an evident increase. The highest average precipitatoin recoreded were all after the year 2000. And on average we can see the increasing trend in the US precipitatoin.

Highest and Lowest States in Increase in Precipitatoin

It is important to see what parts of the country are having the most change in precipitation. That way we can better distinguish the effect of climate change on the type of climate. Since the US has almost all types of climates, we can very accurately get how each region of the world will be effected by climate change based on our results in this section.

We are going to split up every state in the US by the average increase in precipitation every year. We are then going to get the top 15 states that have an increasing trend and also the bottom 15 states. This way we can see if there is any trend in precipitaton in different regions and climate types in the US.

To read more about arrays and dataframes, click here

In [16]:

```
allStatesSlope = []
year = precipitation_df['Year']
# itterate throught every column/state in the dataset and put slope in allStates
for (columnName, columnData) in precipitation_df.iteritems():
  if (columnName != 'Year'):
    stateColumn = precipitation_df[columnName]
    m, b = np.polyfit(year, stateColumn, 1)
    allStatesSlope.append([columnName, m])

# sort array in order of highest precipitation increase first
allStatesSlope.sort(key= lambda x: x[1], reverse=True)
# make dataframe out of top 15 increasing precipitation state
top15Slope = allStatesSlope[:15]
top15Slope = pd.DataFrame(top15Slope)
# make dataframe out of bottom 15 in precipitation state
bottom15Slope = allStatesSlope[-16:-1]
bottom15Slope = bottom15Slope[::-1]
bottom15Slope = pd.DataFrame(bottom15Slope)

print('\nTop 15 increasing states: \n\n', top15Slope, '\n\n')
print('Bottom 15 increasing states: \n\n', bottom15Slope)
```

```
Top 15 increasing states:

                0         1
0    Rhode Island  1.815342
1         Vermont  1.796526
2   New Hampshire  1.766672
3   Massachusetts  1.750540
4       Tennessee  1.706419
5         Indiana  1.581826
6     Connecticut  1.577381
7     Mississippi  1.509398
8        Kentucky  1.508619
```

```
9         Illinois  1.486279
10           Maine  1.289949
11        New York  1.270513
12       Louisiana  1.264596
13         Alabama  1.235008
14        Michigan  1.185461


Bottom 15 increasing states:

                  0         1
0           Arizona -0.122226
1        California -0.027005
2              Utah  0.004480
3           Wyoming  0.051074
4          Colorado  0.093000
5        New Mexico  0.172291
6           Montana  0.208692
7            Nevada  0.232462
8            Oregon  0.250099
9             Idaho  0.253468
10           Alaska  0.263678
11          Georgia  0.302112
12       Washington  0.406590
13   South Carolina  0.536496
14         Nebraska  0.542786
```

Top 15 States Plot

We can use many different types of plots for this section. But barplots are the best option for our purpose. We want to get a visual on the top increasing states in precipitation and compare it with the US average.

Now we are going to do a barplot of the top 15 states to better visualize the different states and they're increase in precipitation. This mighte help us and anyone reading this tutorial to better understand the change in precipitation in different regions of the country.

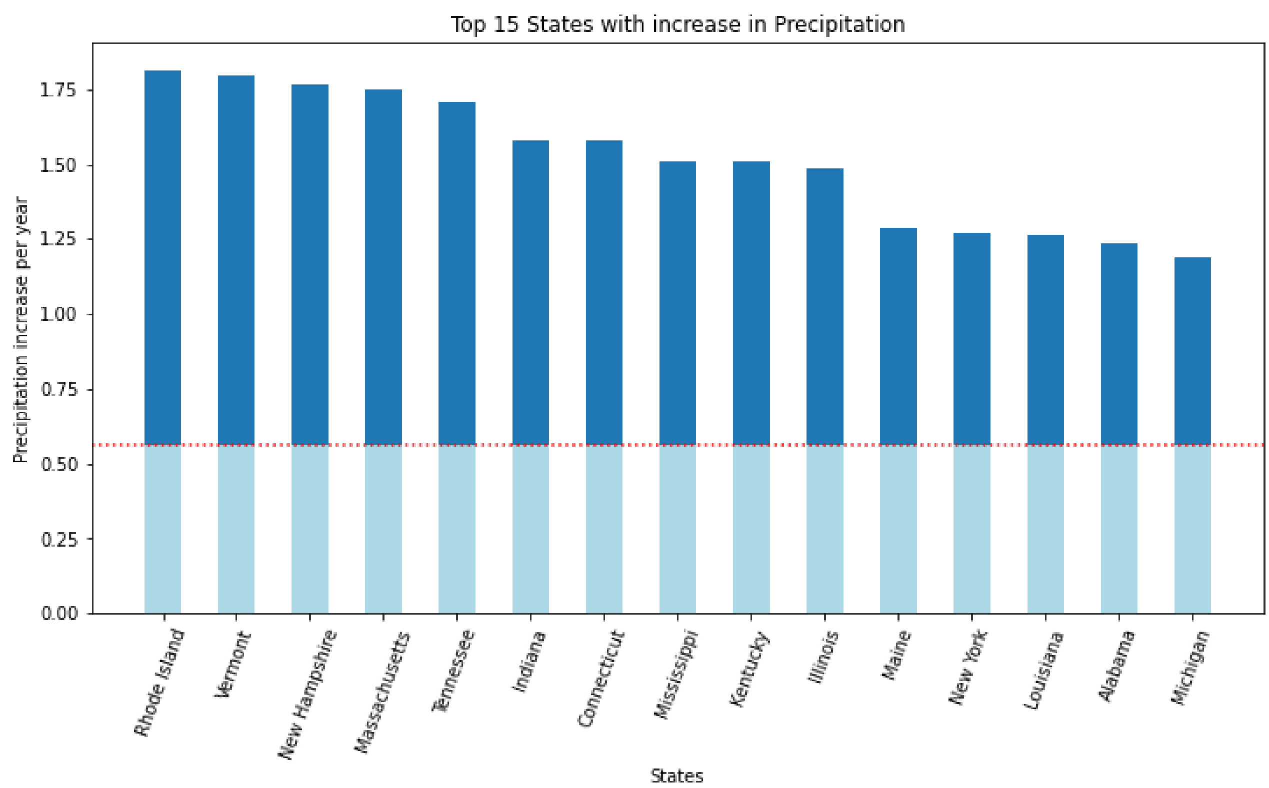
We run a red dotted line indicating the US average.

In [17]:

```
# set the size of the plot to 14 width and 7 height
fig = plt.figure(figsize = (12, 6))

x = precipitation_df['Year']
y = precipitation_df['United States']
#obtain m (slope) and b(intercept) of linear regression lineir
m, b = np.polyfit(x, y, 1)

usSlope = m

# upper half of the regional bar. Remainder of precip increase passed the US ave
athreshold = list(map(lambda a : a - usSlope if a - usSlope > 0 else 0, top15Slo

# creating the bar plot
# lower half colored green
plt.bar(top15Slope[0], top15Slope[1], color ='#ADD8E6',
      width = 0.5)
# upper half colored red
plt.bar(top15Slope[0], athreshold, width = 0.5,bottom=usSlope)
#xticks
plt.xticks(rotation=70)

#x-axis labels
plt.xlabel('States')

#y-axis labels
plt.ylabel('Precipitation increase per year')

#plot title
plt.title('Top 15 States with increase in Precipitation')

# put dotted line indicating US average
plt.axhline(usSlope, color='red', ls='dotted')
```

```
<matplotlib.lines.Line2D at 0x7fa2406488e0>
```

Out[17]:



Top 15 Increasing States Analysis

In this barplot we can see that these states have more than twice the US avergae precipitation increase. That is a staggering amount. Now, we're going to see if there is any pattern in the regions in these top 15 states.

We can see that the top 15 states with the highest increase in percipitation are mostly from North Eastern Part of the US. Rhode Island, Vermont, New Hampshire and Massachusetts are just the top 4 states and they are all in the most North East part of the US. There are also a few states more towards the Central US, but still East such as Michigan and Mississippi. However in the top 15, we do not have a single state from the West. And they are mostly states towards the North. We can see from this that precipitation is being effected more significantly in the North and places where precipitation was already high.

Bottom 15 States Plot

As described above, we can use many different types of plots for this section. But barplots are the best option for our purpose. We want to get a visual on the bottome decreasing/increasing states in precipitation and compare it with the US average.

Now we are going to do a barplot of the bottom 15 states to better visualize the different states and they're decrease/increase in precipitation.

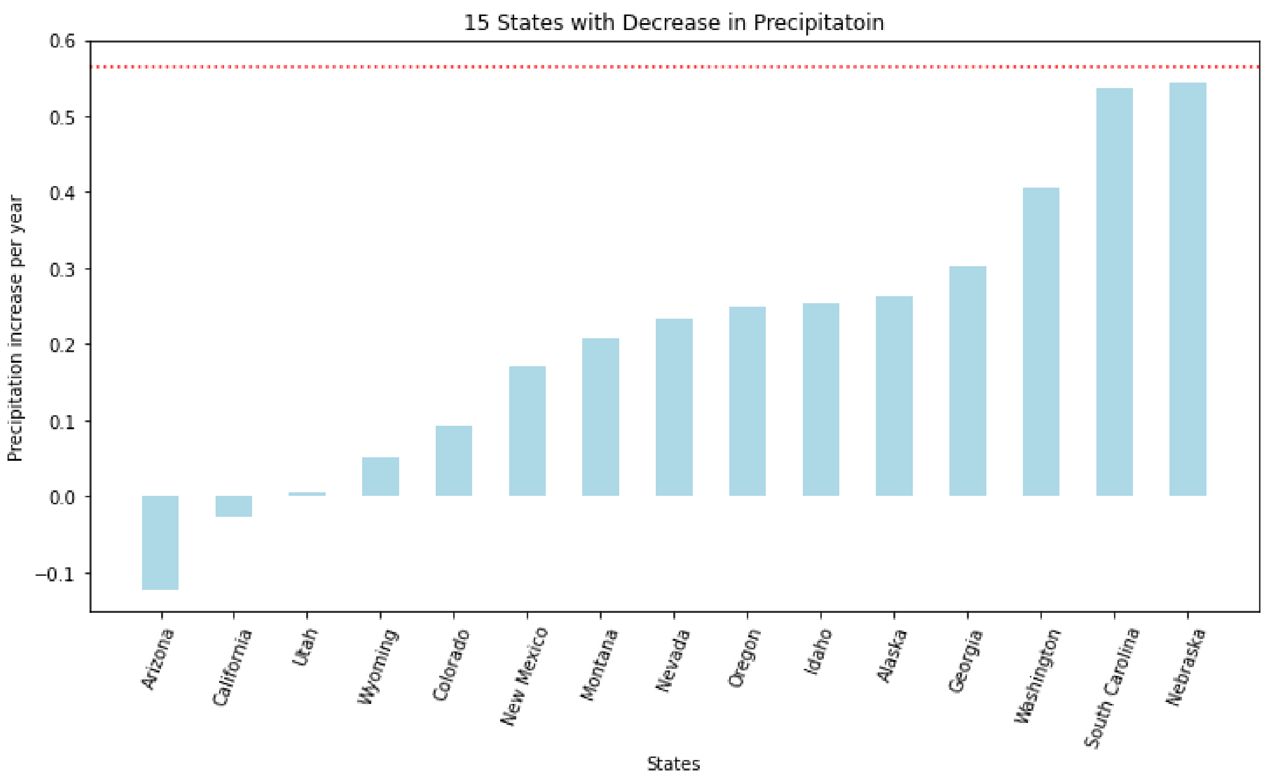
We run a red dotted line indicating the US average.

In [18]:

```
# set the size of the plot to 14 width and 7 height
fig = plt.figure(figsize = (12, 6))

x = precipitation_df['Year']
y = precipitation_df['United States']
#obtain m (slope) and b(intercept) of linear regression lineir
m, b = np.polyfit(x, y, 1)

usSlope = m

# upper half of the regional bar. Remainder of precip increase passed the US ave
athreshold = list(map(lambda a : a - usSlope if a - usSlope > 0 else 0, bottom15

# creating the bar plot
# lower half colored green
plt.bar(bottom15Slope[0], bottom15Slope[1], color ='#ADD8E6',
        width = 0.5)
```

```
# upper half colored red
plt.bar(bottom15Slope[0], athreshold, width = 0.5,bottom=usSlope)
#xticks
plt.xticks(rotation=70)

#x-axis labels
plt.xlabel('States')

#y-axis labels
plt.ylabel('Precipitation increase per year')

#plot title
plt.title('15 States with Decrease in Precipitatoin')

# put dotted line indicating US average
plt.axhline(usSlope, color='red', ls='dotted')
plt.ylim([-0.15,0.6])
```

Out[18]: (-0.15, 0.6)



Bottom 15 States Analysis

In contrast to the top 15 states, we can see that all of the states that either have a decrease in precipitation or have a very slight increase in precipitation are Central and Western states. We can see that Arizona and California have had a decrease in average precipitation. This can be because, most of these states are drier and thus the climate change has not effected them as much.

Precipitation Conclusion

We can conclude that overall, the US annual precipitation is increasing. Even in the bottom 15 states, only two of them had a decreasing trend in their precipitation and the rest were all increasing but just at a lower rate.

There is also a strong link between the region in the country and the amount that precipitation is increasing. We might be able to find a correlation between the region and the temperature change and that is what we are going to find out in the next section.

Correlation Between Temperature Change and Precipitation in the US

Finally we need to see if there is any correlation between the precipitation increase and the temperature change. As seen before, there is an increase in temperature across the globe and especially in the North America regions. We also saw that we are having increase in precipitation in most parts of the US. Here we are going to see if we can find any correlation between the two.

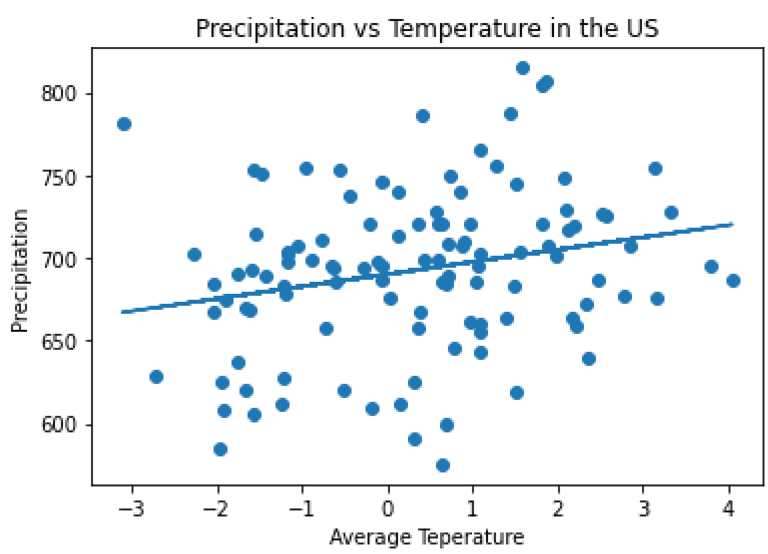
The best way to show this correlation is by a simple scatter plot, because we can see how correlated or evenly distributed the plot is by simply looking at it.

We are first going to make a scatter plot of the Precipitatoin vs Temperature in the US to find a trend. The trend should be an increasing slope, since both temperature and precipitation are increasing together.

In [19]:

```
# set x to temperature in North America
x = climate_dfs[5][:-2]['jan_temp_celsius']
# set y to precipitation in America
y = precipitation_df[9:]['United States']
plt.scatter(x=x, y=y)
plt.xlabel('Average Teperature')
plt.ylabel('Precipitation')
plt.title('Precipitation vs Temperature in the US')
#obtain m (slope) and b(intercept) of linear regression lineir
m, b = np.polyfit(x, y, 1)

#add linear regression line to scatterplot
plt.plot(x, m*x+b)
```

Out[19]: [<matplotlib.lines.Line2D at 0x7fa2004bff10>]



There is not much correlation that we can see from this graph. However, after running a regression line through it, we can see a positive correlation. However, this is not convincing enough so we will have to do some more analysis.

We are going to see if there is any upward trend between precipitation and temperature in the top increasing states (in terms of precipitation) and if there is any downward trend between precipitation and temperature in the bottom states (in terms of precipitation)

Top Increasing Precipitation vs Temperature

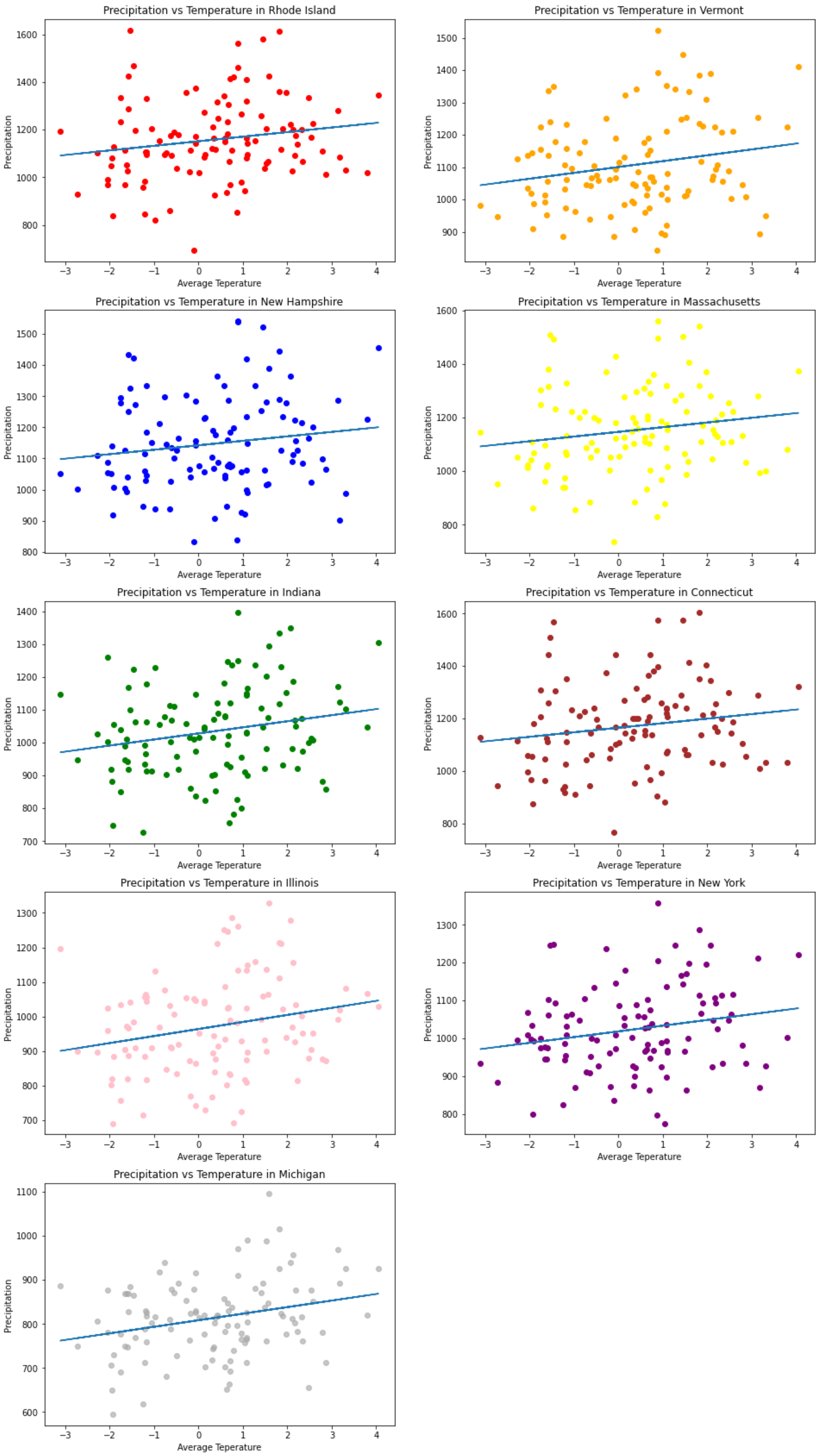
Here we are going to plot the top states that have shown highest increase in precipitation in the US and we are going to plot them agains temperature. By doing this we can see if there is any trend.

If you want to learn more about subplots click Subplots.

In [20]:

```
# set the size of the plot to 14 width and 7 height
fig = plt.figure(figsize = (16, 30))

topStates = ['Rhode Island', 'Vermont', 'New Hampshire', 'Massachusetts', 'India
colors = ['red', 'orange', 'blue', 'yellow', 'green', 'brown', 'pink', 'purple',
# set x to temperature in North America
```

```
x = climate_dfs[5][:-2]['jan_temp_celsius']

i = 1
# go through all top states and plot precipitation vs temp for the 9 selected st
for state in topStates:
  # creating subplots to put side by side
  plt.subplot(5, 2, i) # row 5, col 2 index i
  y = precipitation_df[9:][state]
  plt.scatter(x=x, y=y, color=colors[i-1])
  plt.xlabel('Average Teperature')
  plt.ylabel('Precipitation')
  plt.title('Precipitation vs Temperature in {}'.format(state))
  #obtain m (slope) and b(intercept) of linear regression lineir
  m, b = np.polyfit(x, y, 1)
  #add linear regression line to scatterplot
  plt.plot(x, m*x+b)
  i += 1
plt.show()
```



We can see that some of the states have a very good correlation between their increase in precipitation and the temperature. Michigan, New York, Illinois and Indiana show a definite increase as temperature increases. However, states such as Connecticut, Massachusetts and New Hamshire do not show much correlation and their plots are very scattered meaning that even at lower temperatures we have gotten high precipitation readings.

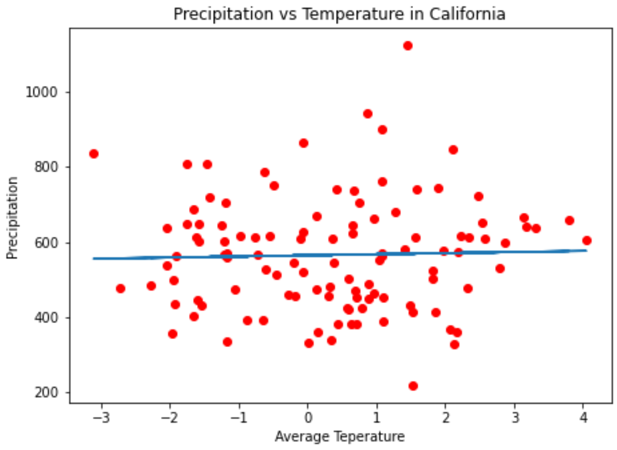
One thing that however, they all have in common is that all of their regression lines are sloped upwards. Meaning that they all have at least a slight increase in precipitation when temperature increases.

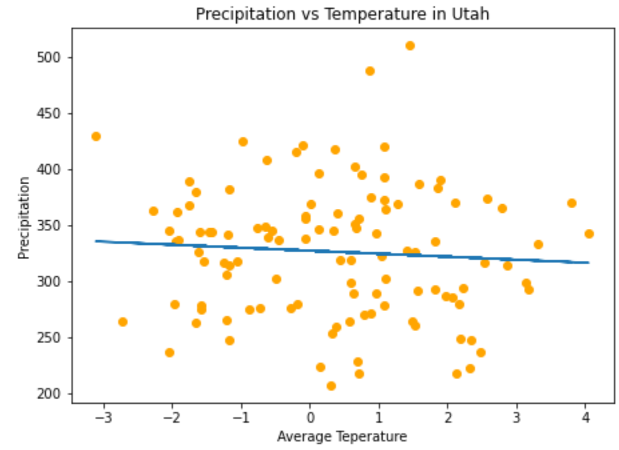
Most Decrease/Least Increase Precipitation vs Temperature

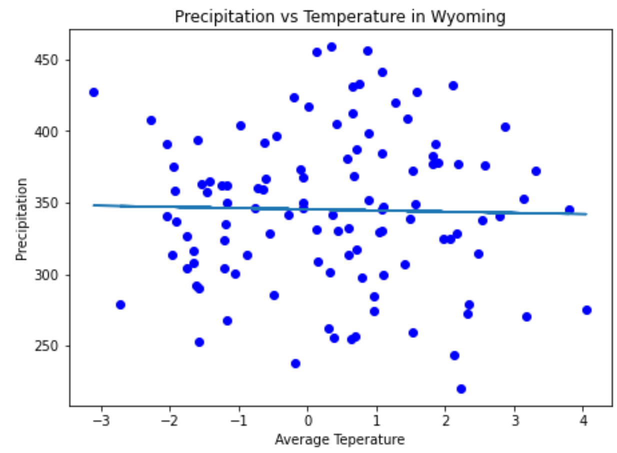
Here we are going to plot the bottom states that have shown the least increase or even a decrease in precipitation in the US and we are going to plot them agains temperature. By doing this we can see if there is any trend.

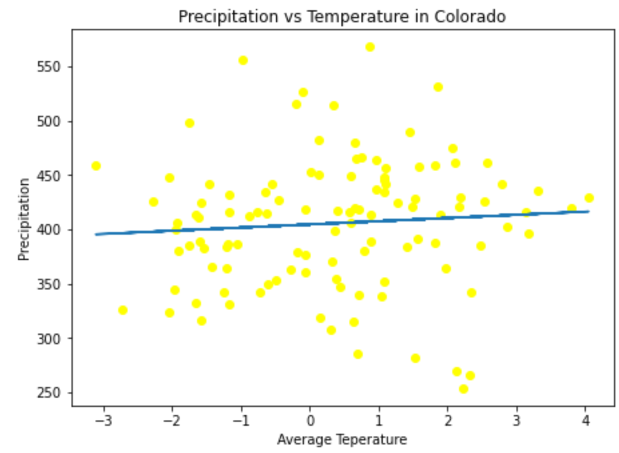
In [21]:

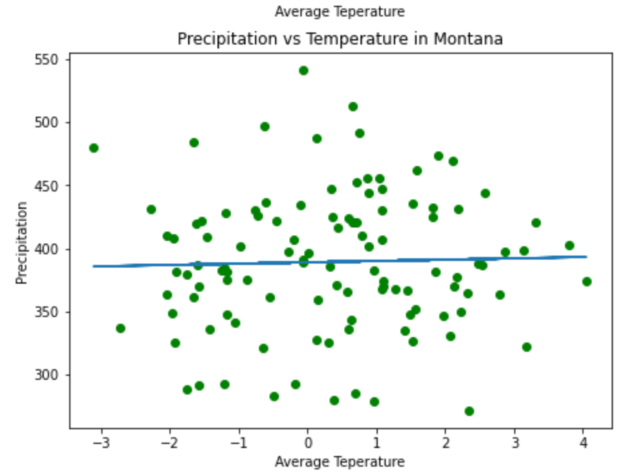
```
# set the size of the plot to 14 width and 7 height
fig = plt.figure(figsize = (16, 30))

bottomStates = ['California', 'Utah', 'Wyoming', 'Colorado', 'Montana', 'Nevada'
colors = ['red', 'orange', 'blue', 'yellow', 'green', 'brown', 'pink', 'purple',
# set x to temperature in North America
x = climate_dfs[5][:-2]['jan_temp_celsius']

i = 1
# go through all bottom states and plot precipitation vs temp for the 9 selected
for state in bottomStates:
  # creating subplots to put side by side
  plt.subplot(5, 2, i) # row 5, col 2 index i
  y = precipitation_df[9:][state]
  plt.scatter(x=x, y=y, color=colors[i-1])
  plt.xlabel('Average Teperature')
  plt.ylabel('Precipitation')
  plt.title('Precipitation vs Temperature in {}'.format(state))
  #obtain m (slope) and b(intercept) of linear regression lineir
  m, b = np.polyfit(x, y, 1)
  #add linear regression line to scatterplot
  plt.plot(x, m*x+b)
  i += 1
plt.show()
```

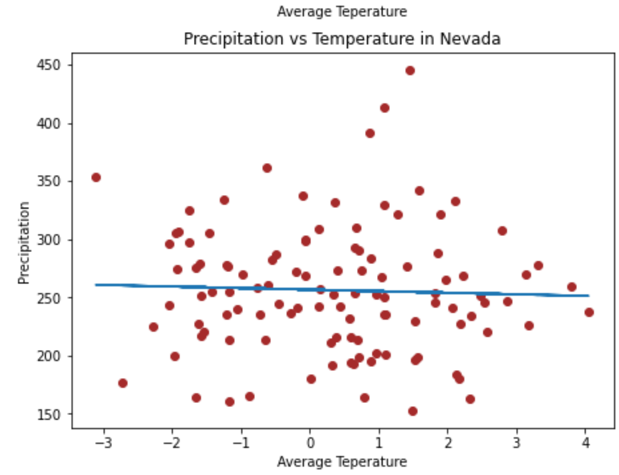


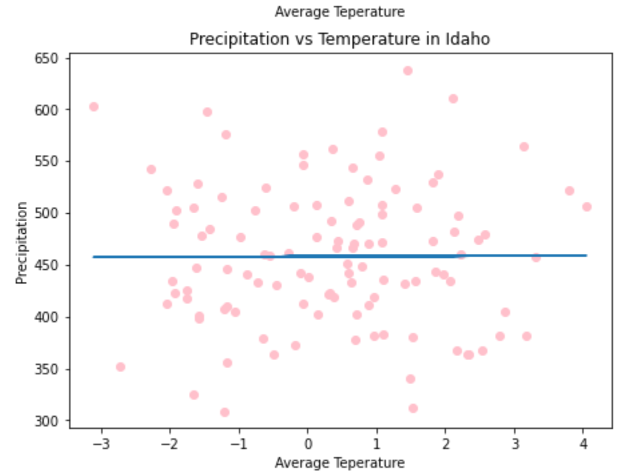


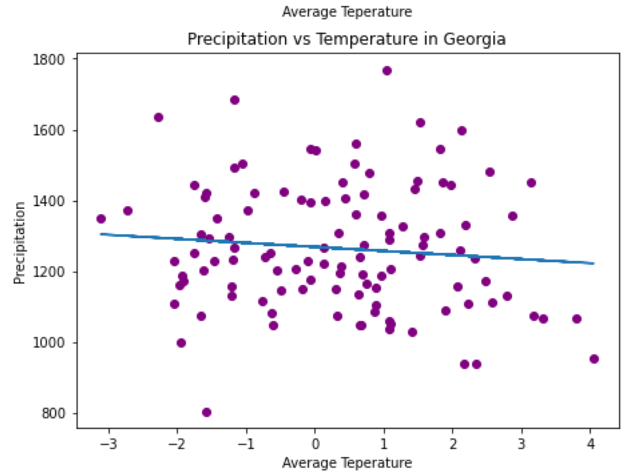


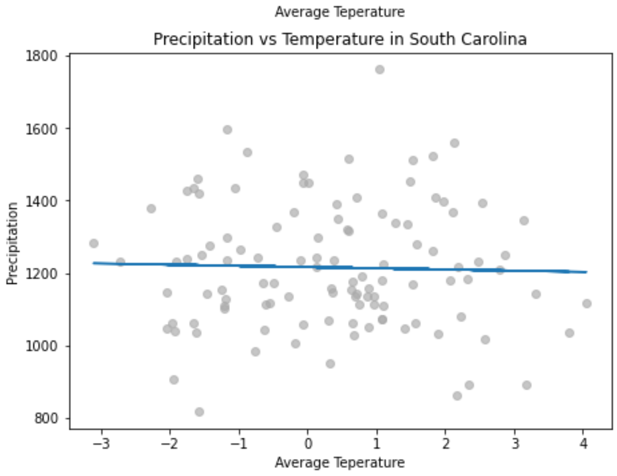












In this part we have gotten some very interesting data. We have a very flat slope for most of the states, meaning that as temperature increases, the precipitation pretty much stays the same. In some states we are even getting a negative slope. However, this makes sene, since these states we had already established that they didn't have much change in precipitation and two of them had a decrease in precipitation.

Since, these states were mostly in the wester/central parts of the US and also these are the more dry states in the US. We can conclude that the temperature is not effecting drier regions of the country as much as it is effecting states that already had high precipitations.

Conclusion

In this tutorial we have went over so many different aspects of climate change and precipitatoin. We have seen that climate change is not effecting every region of the world at the same rate. The degree of change depends on so many variables. The variables that we were able see was that, at regions closer to the equator, we had less change in the temperature. In contrast we had the most change in North America which is above the equator.

We then analysed the precipitation patterns in the United States. We saw that precipitation was generally increasing the US. We then analysed the states individually and found out that the most drastic changes in temperature are happening in the North Eastern parts of the US. And in contrast the least effected states are states in the West coast or in the Central US. This comes to show again that as the temperatures are rising, the effects are not going to be evenly distributed and there are biases based on the regions. The correlation between increase in Temperature and Precipitation are more evident in states that have higher overall precipitation.

In this tutorial, we have used a series of plots and regression lines. Barplot and scatter plots were the two different types of plots we used. We also used regression lines and average line to make better sense of our data. For more resources and tutorials on what we used, we have included some links in the "Refernces" section.

References:

- Preciptitaion Data from the United States Environmental Protection Agency - https://www.epa.gov/climate-indicators/climate-change-indicators-us-and-global-precipitation
- Annual Global Temperature Averages - https://datahub.io/core/global-temp#data
- National Centers for Environmental Information, National Oceanic and Atmospheric Administration (NOAA) - https://www.ncdc.noaa.gov/cag/global/time-series/
- Scatter Plots - https://matplotlib.org/3.5.0/api/_as_gen/matplotlib.pyplot.scatter.html
- Linear Regression - https://realpython.com/linear-regression-in-python/
- *Which parts of the planet are warming the fastest, and why?* - https://climate.mit.edu/ask-mit/which-parts-planet-are-warming-fastest-and-why#:~:text=The%20earth's%20largest%20land%20masses,reflect%20energy%20from%20th
- Barplot - https://matplotlib.org/3.5.0/api/_as_gen/matplotlib.pyplot.bar.html
- Dataframes and Arrays - https://www.geeksforgeeks.org/create-a-dataframe-from-a-numpy-array-and-specify-the-index-column-and-column-headers/
- Subplots - https://www.tutorialspoint.com/how-to-make-two-plots-side-by-side-using-python