****Московский государственный технический университет им. Н.Э. Баумана  
Кафедра «Системы обработки информации и управления»

Лабораторная работа №1  
по дисциплине  
«Методы машинного обучения»  
на тему

## «Создание "истории о данных" (Data Storytelling)»

Выполнил:  
студент группы ИУ5-21М  
Чжан Чжибо

Москва — 2021 г.

**1. Цель лабораторной работы**

Изучение различных методов визуализация данных и создание истории на основе данных.

**2. Задание**

2.1 Выбрать набор данных (датасет).

2.2 Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:

История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.

На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.

Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.

Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.

История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.

2.3 Сформировать отчет и разместить его в своем репозитории на github.

**3. Ход выполнения работы  
3.1. Текстовое описание набора данных**

Глобальное потепление-это феномен повышения температуры Земли, воздуха и океана за последние пару столетий. Для этого состояния характерна человеческая деятельность, в первую очередь сжигание ископаемого топлива.

В работе использованы наборы данных GlobalTemperatures.csv, GlobalLandTemperatureByCountry.csv, GlobalLandTemperatureByStates.csv из Kaggle, предоставленного пакетом данных Berkeley Earth.

Цель работы: проверка изменения температуры по:

1) годам

2) месяцам

3) временам года

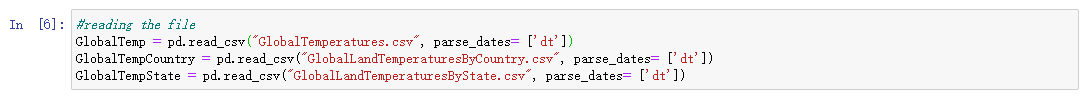
4) странам

Самое главное это проверить, правда ли, что температура становится выше.

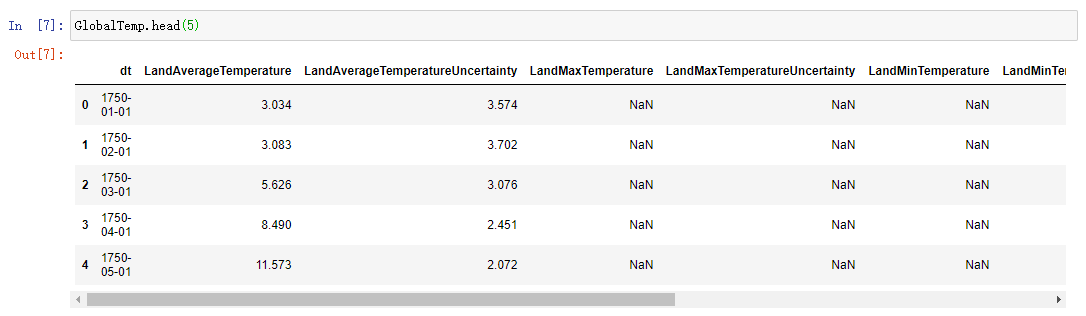
**3.2. Основные характеристики набора данных**  
Подключим все необходимые библиотеки:



Загрузим непосредственно данные:



Показание информаций о температуре суши и океана на земле за последние несколько лет:

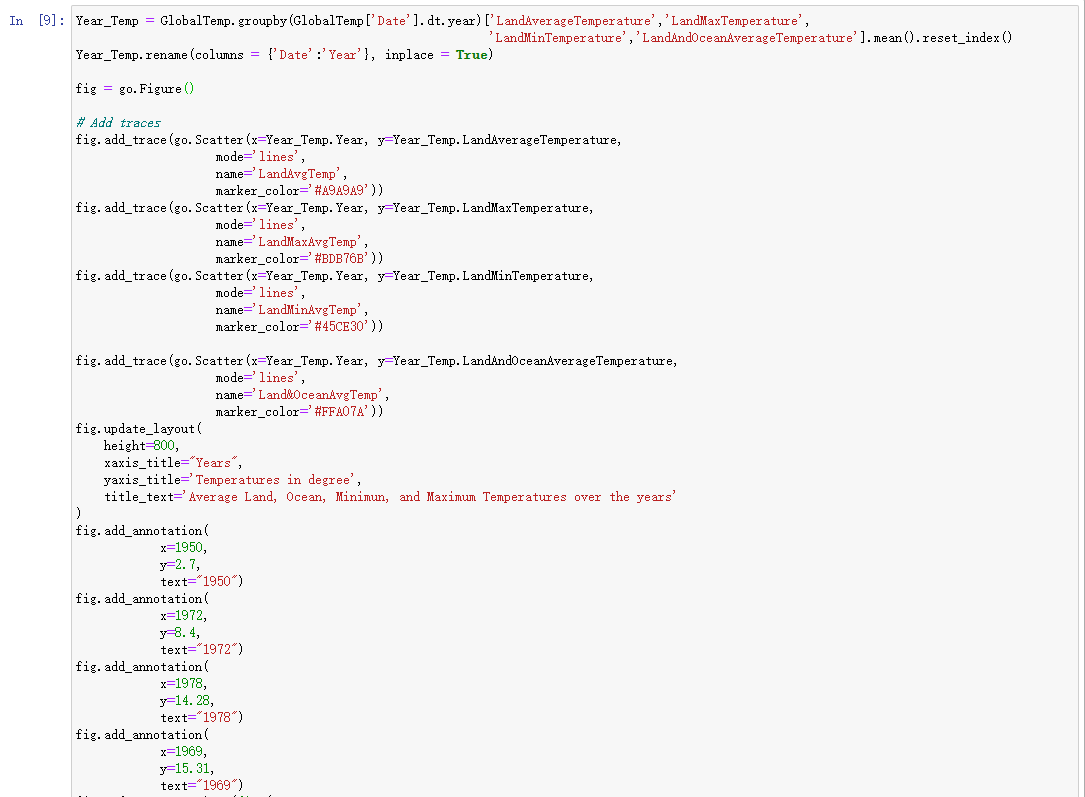


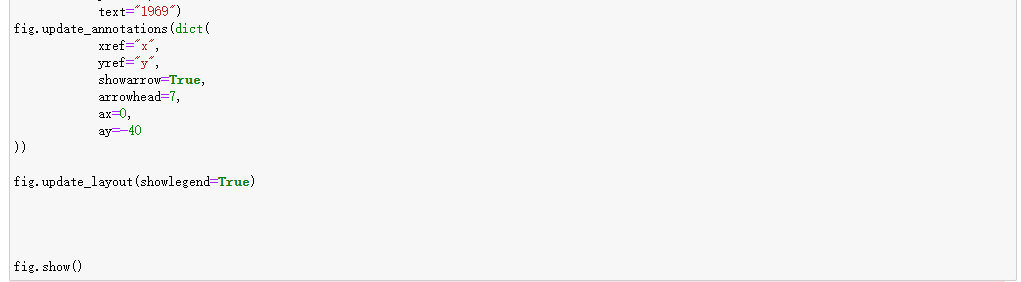
Изменение имени столбца для удобства:

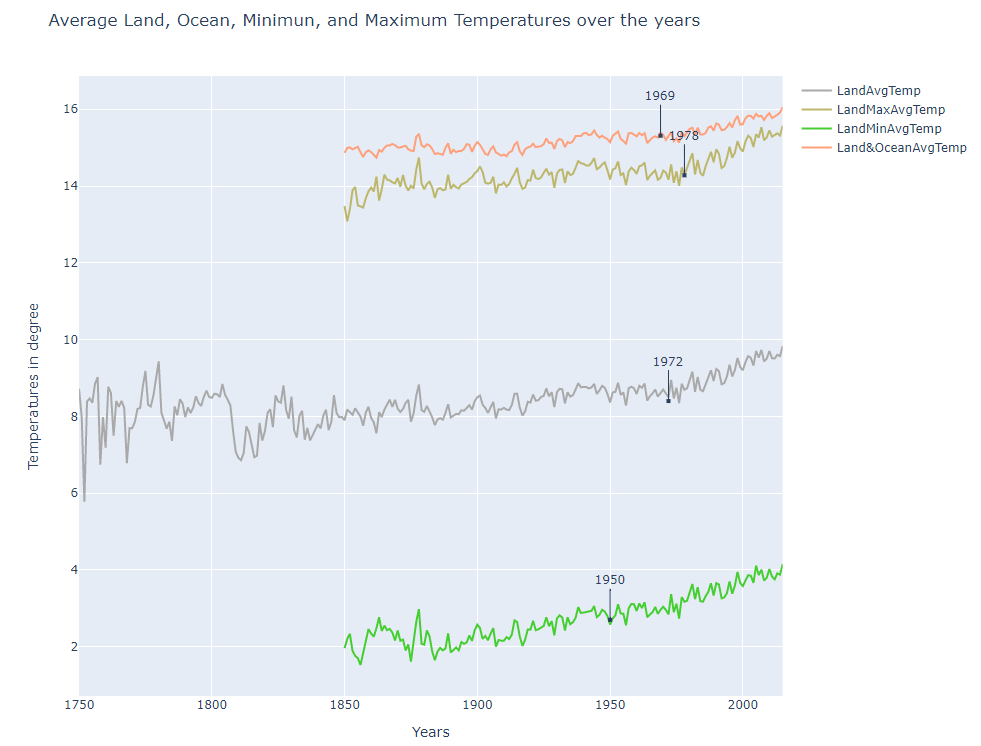


**3.3. Визуальное исследование датасета**

Найдём время, когда температура начала расти.







Похоже, что данные по LandMaxTemperature, LandMinTemperature, Land&OceanAverageTemperature отсутствуют с 1750 по 1850 год.

Очевидно, что:

Средняя температура земли и океана увеличилась на 1,19 градуса с 1850 по 2015 год.

Максимальная температура земли увеличилась на 2,1 градуса с 1850 по 2015 год.

Средняя температура земли увеличилась на 1,12 градуса с 1750 по 2015 год

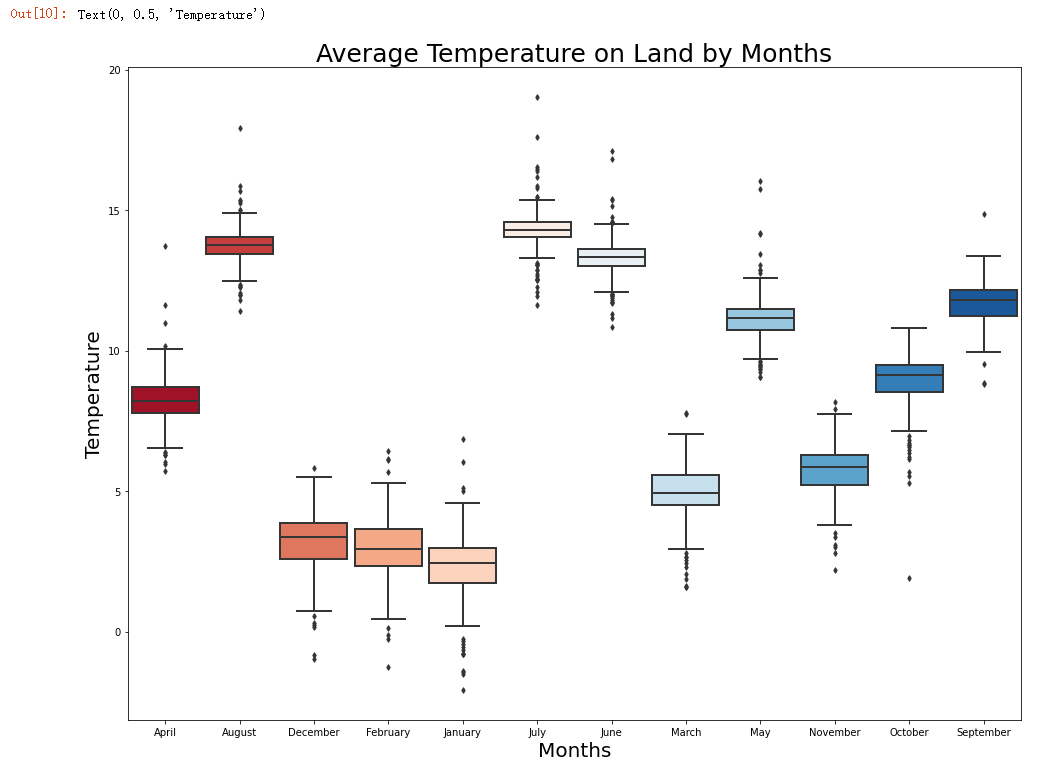
Минимальная температура земли достигла максимума в 2,24 градуса с 1850 по 2015 год

Из графика видно, что начиная с 1950 года все параметры стали увеличиваться.

Теперь двинемся вперед и проведем ежемесячный анализ, чтобы увидеть, какие месяцы удерживают высокие температуры за эти годы

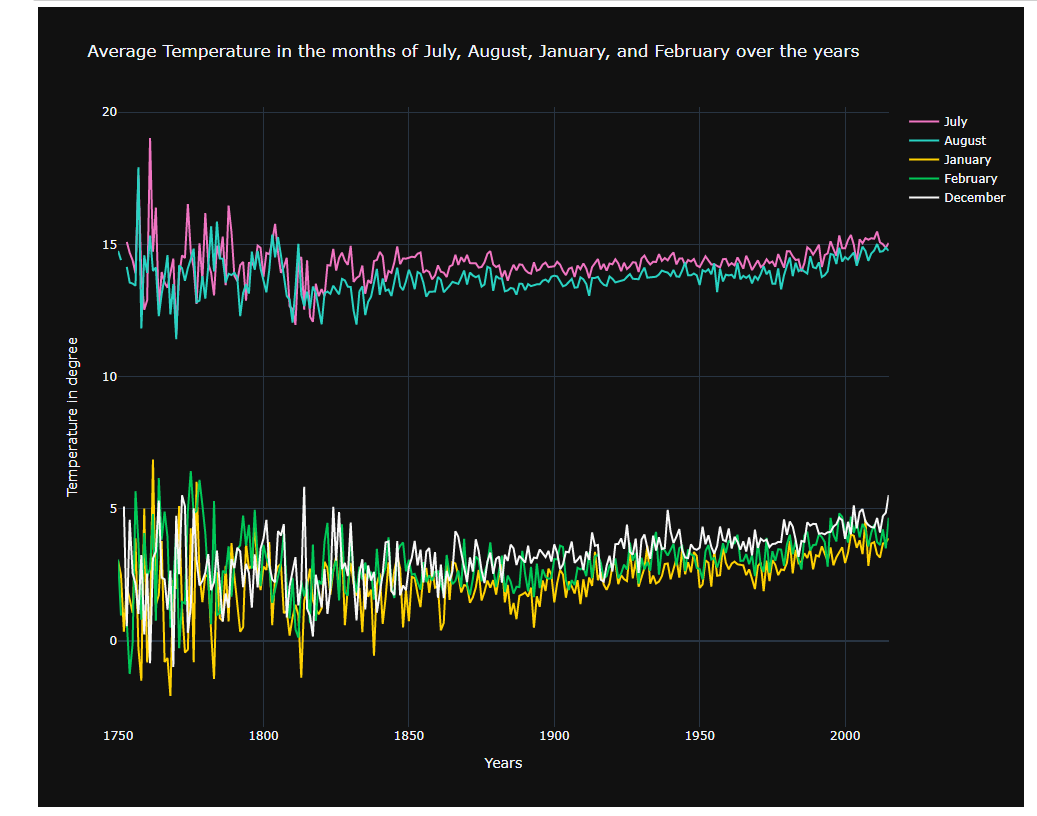
Ежемесячный анализ температуры по годам:

****

****

Как мы знаем, температура,по-видимому,была высокой в июле и августе и низкой в январе,феврале и декабре на протяжении 250 лет. Дальше копнем глубже и увидим повышение температуры с годами за эти месяцы





Видно, что:

Средняя температура поднялась на 0,041 градуса в июле месяце с 1753 по 2015 год.

Средняя температура поднялась на 0,005 градуса в августе месяце с 1750 по 2015 год.

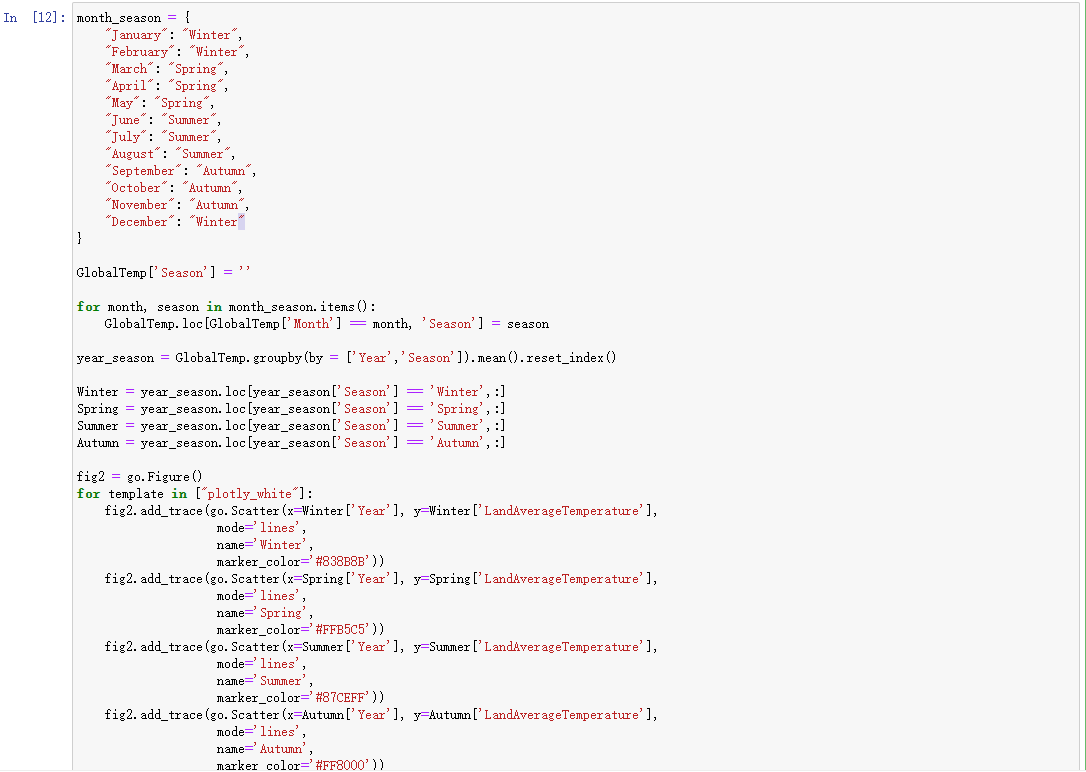
Средняя температура поднялась почти на 1 градус в январе с 1750 по 2015 год.

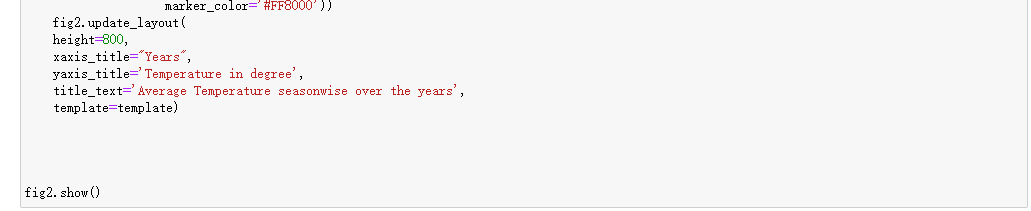
Средняя температура поднялась на 1,58 градуса в феврале месяце с 1750 по 2015 год.

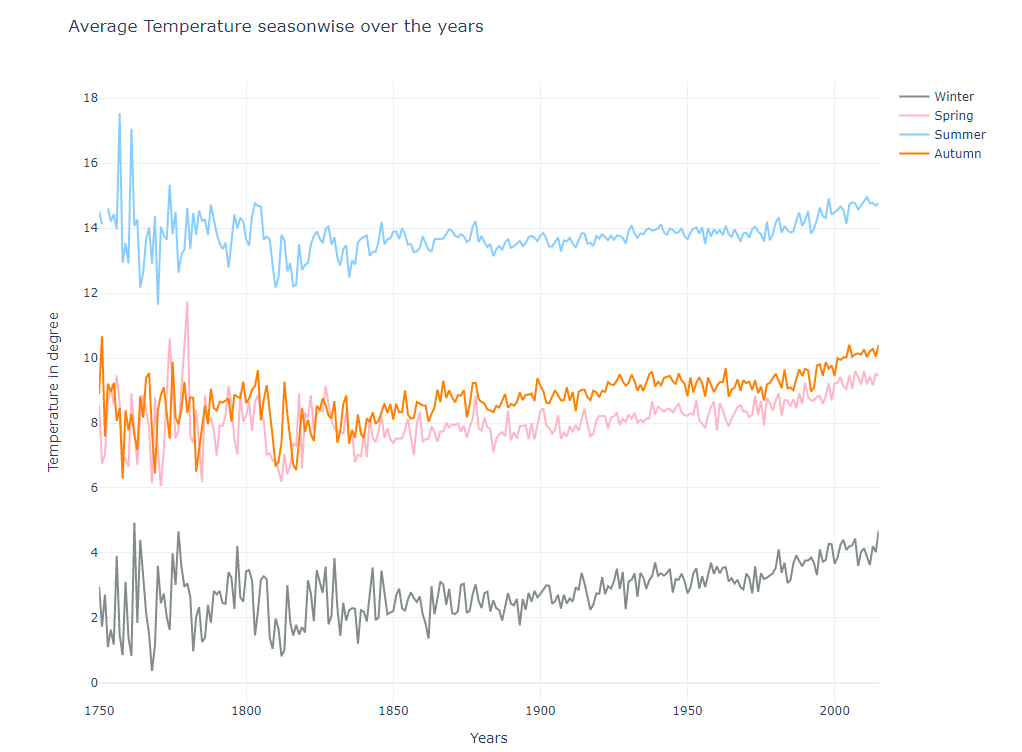
Средняя температура поднялась на 2,746 градуса в декабре с 1750 по 2015 год.

По разнице температур вышеописанных месяцев можно сказать, что в холодные месяцы разница температур выше, чем в жаркие. Это признак того, что холодные месяцы становятся все жарче.

Теперь попроверим, не становится ли какое-нибудь время года жарче с годами. Посмотрим сезонное повышение температуры за данные годы. Поскольку мы имеем дело с данными временных рядов, будет предпочтительнее снова использовать линейный график

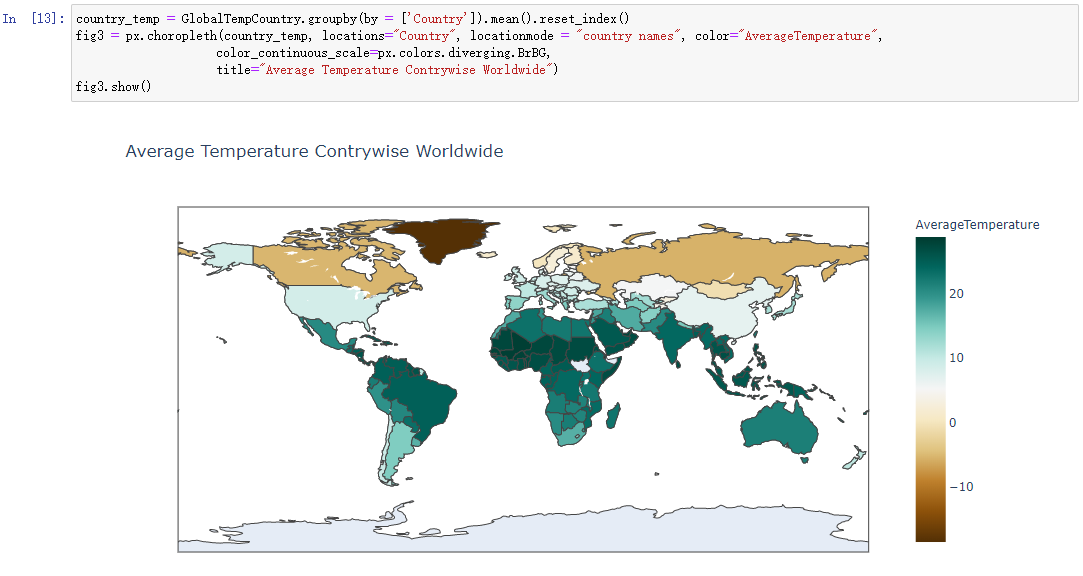






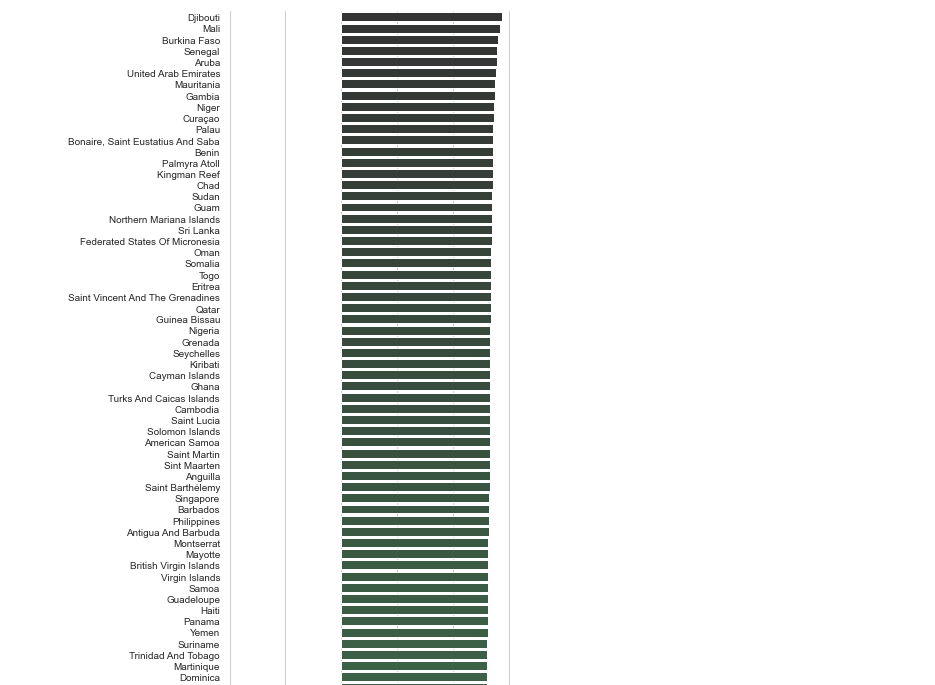
Из графика видно, что весна, лето и зима с каждым годом становятся все жарче. Особенно зимний сезон имеет самый высокий всплеск за последние годы. Таким образом, мы можем сказать, что холодные сезоны становятся все жарче.

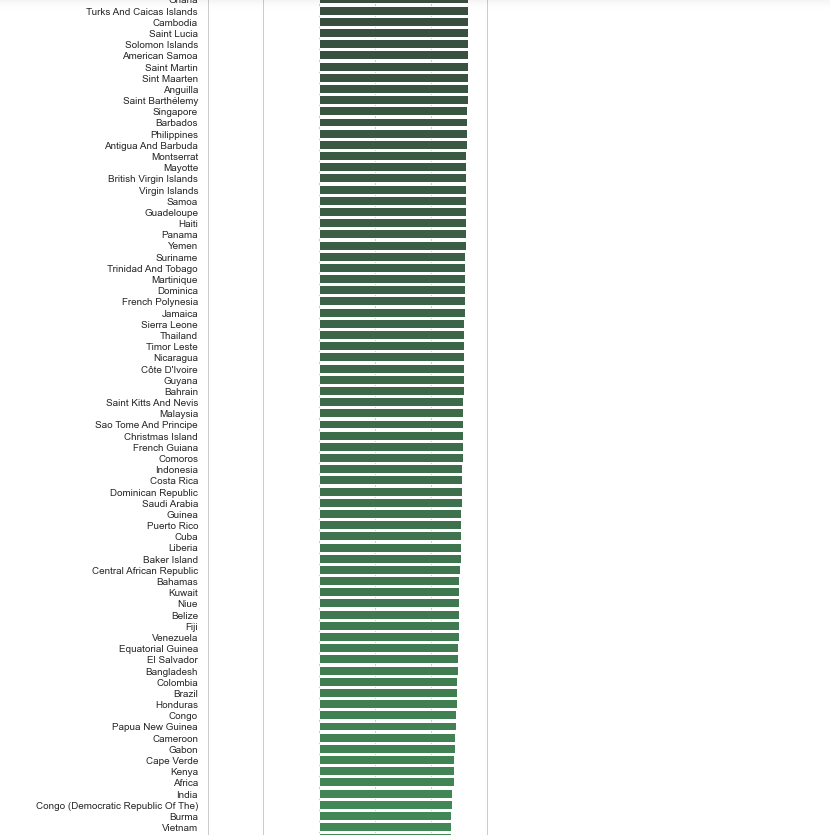
Перейдем к другому датсету и проверим пострановой средней температуры страны за эти годы:

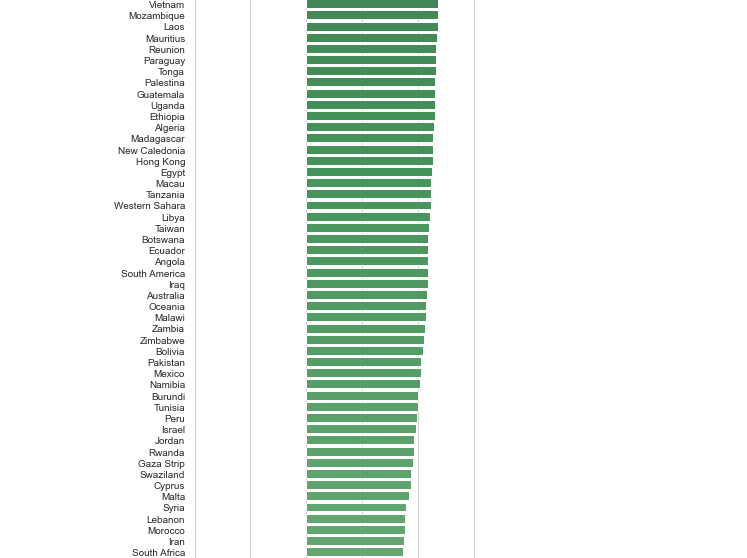


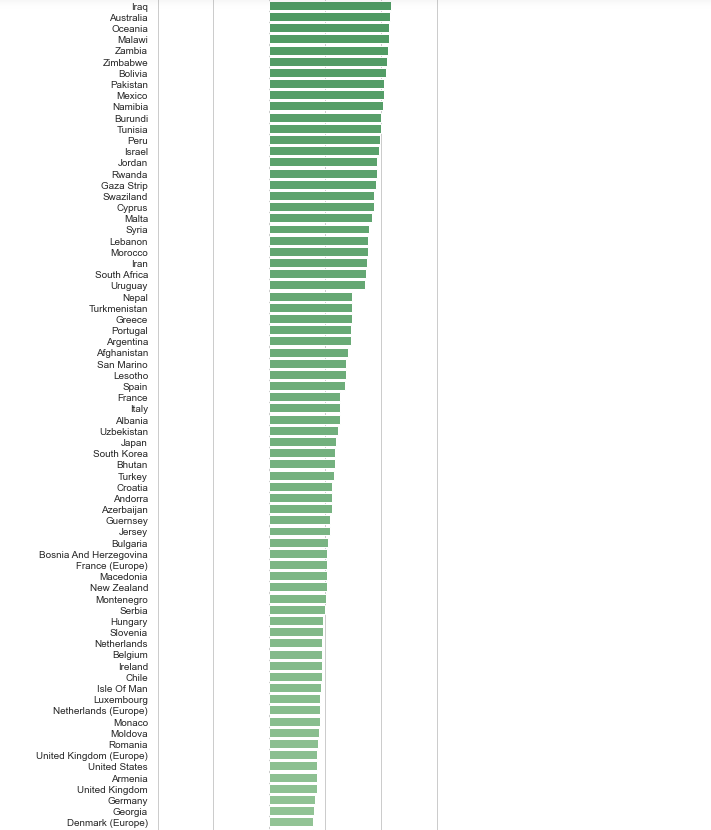
Из приведенного выше хороплета видно,что страны Гренландия, Россия, Канада имеют самую низкую среднюю температуру,а страны африканского континента, такие как Джибути, Мали, Буркина-Фасо, имеют самую высокую среднюю температуру в течение многих лет.

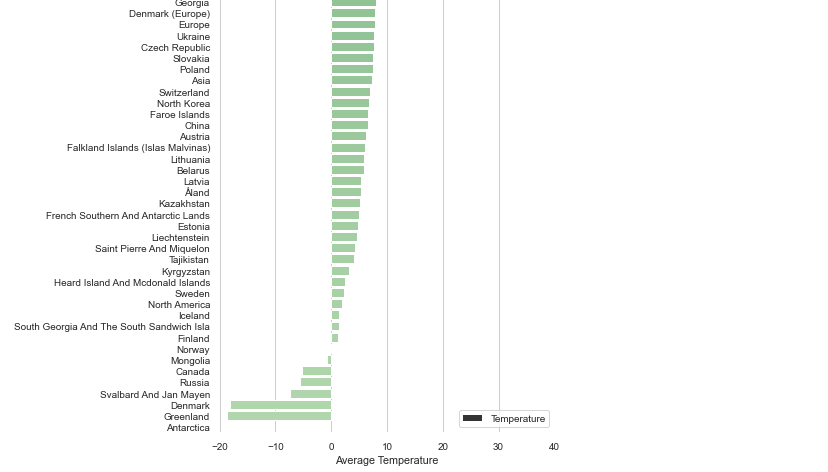


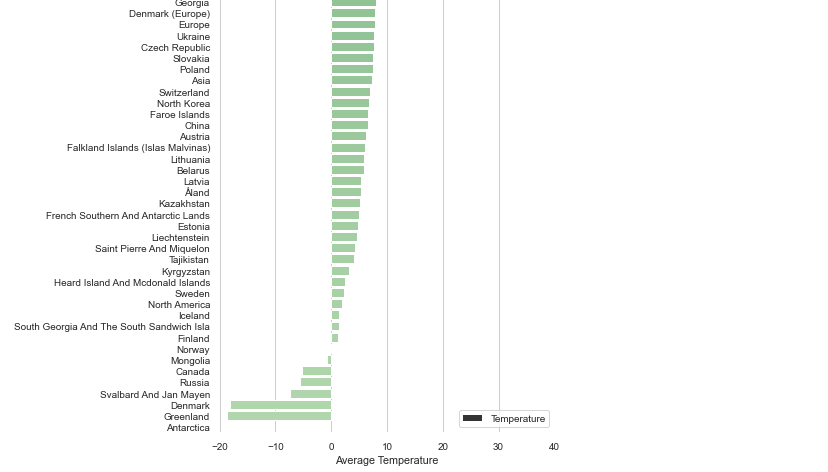




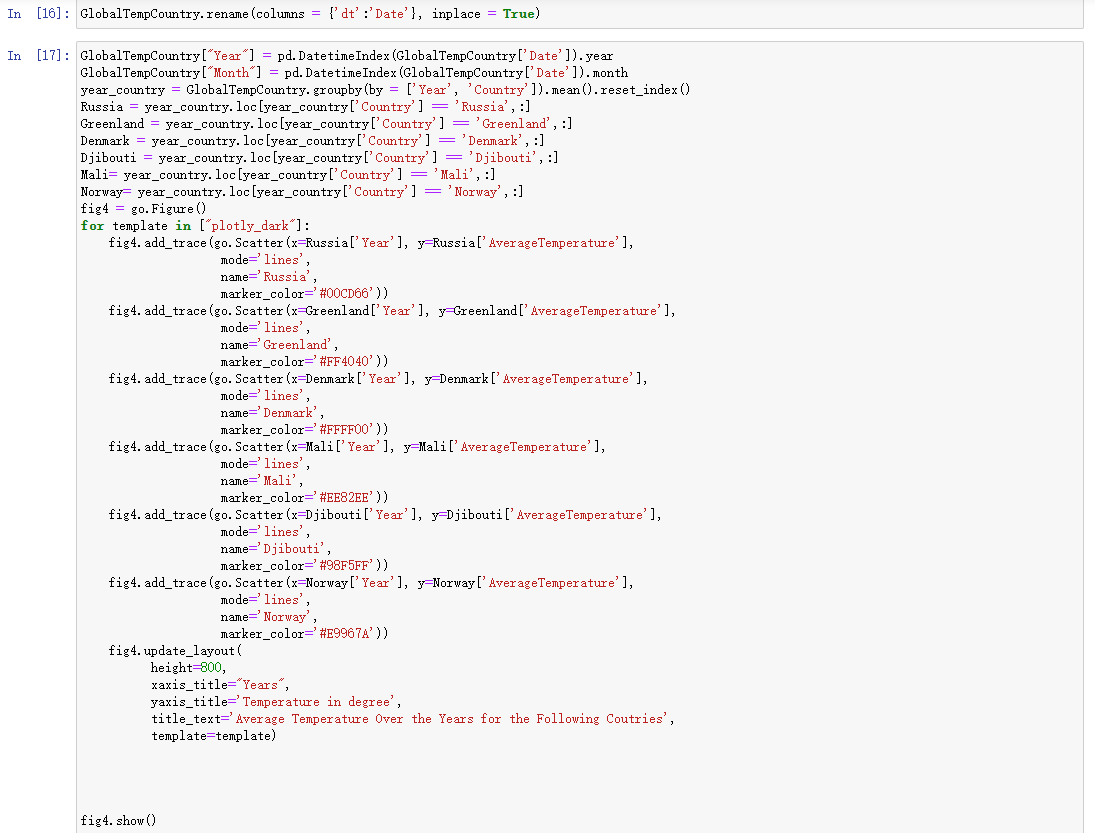


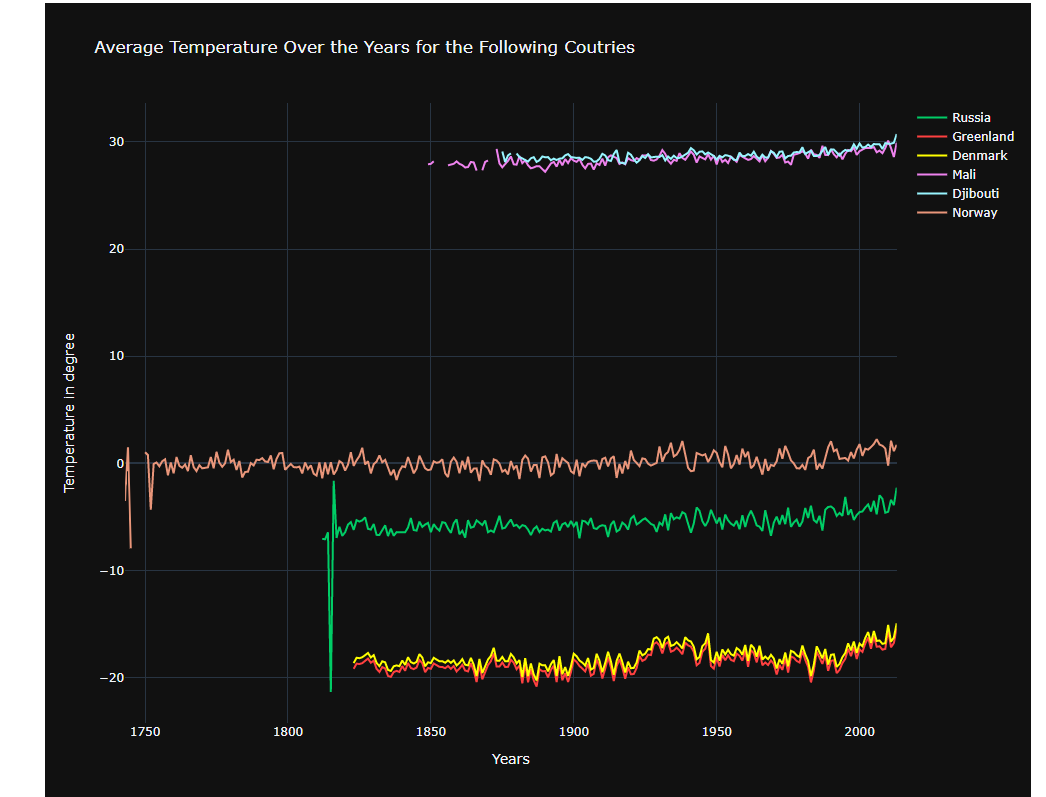






Теперь анализируем некоторые из самых жарких и прохладных стран и посмотрим, наблюдается ли повышение температуры с годами. Я хотел бы проанализировать Гренландию, Данию, Россию, Джибути и Мали





Из графика видно, что:

Гренландия имеет разницу в 3,77 градуса за данные годы.

Денамрк имеет разницу в 3,7 градуса за данные годы.

В России за эти годы разница составляет 4,78 градуса.

Норвегия имеет разницу в 5,2 градуса за данные годы.

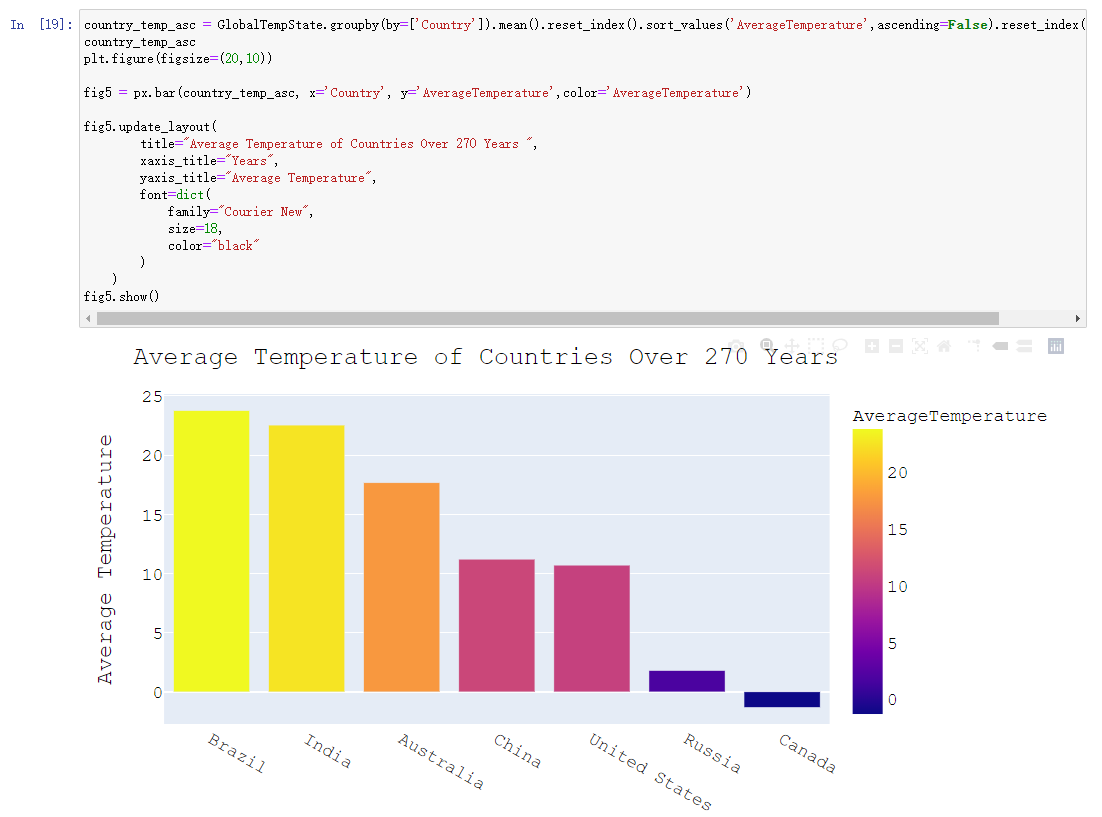
В то время страны с высокими температурами:

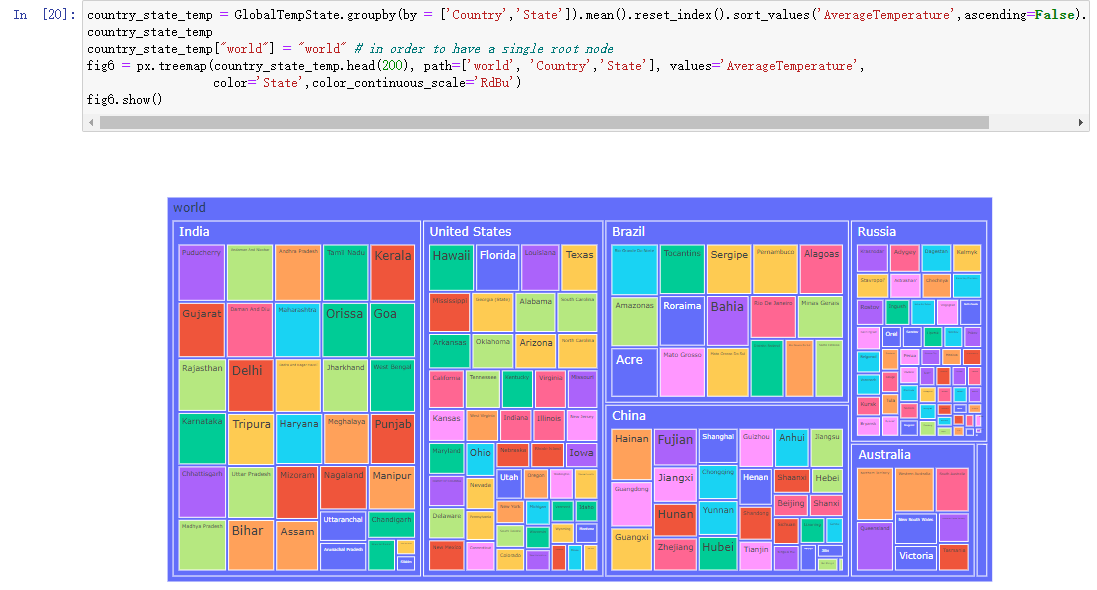
Мали имеет разницу в 1,99 градуса за данные годы.

Джибути имеет разницу в 0,89 градуса за данные годы.

Таким образом, мы можем определенно сказать, что холодные места становятся все жарче.

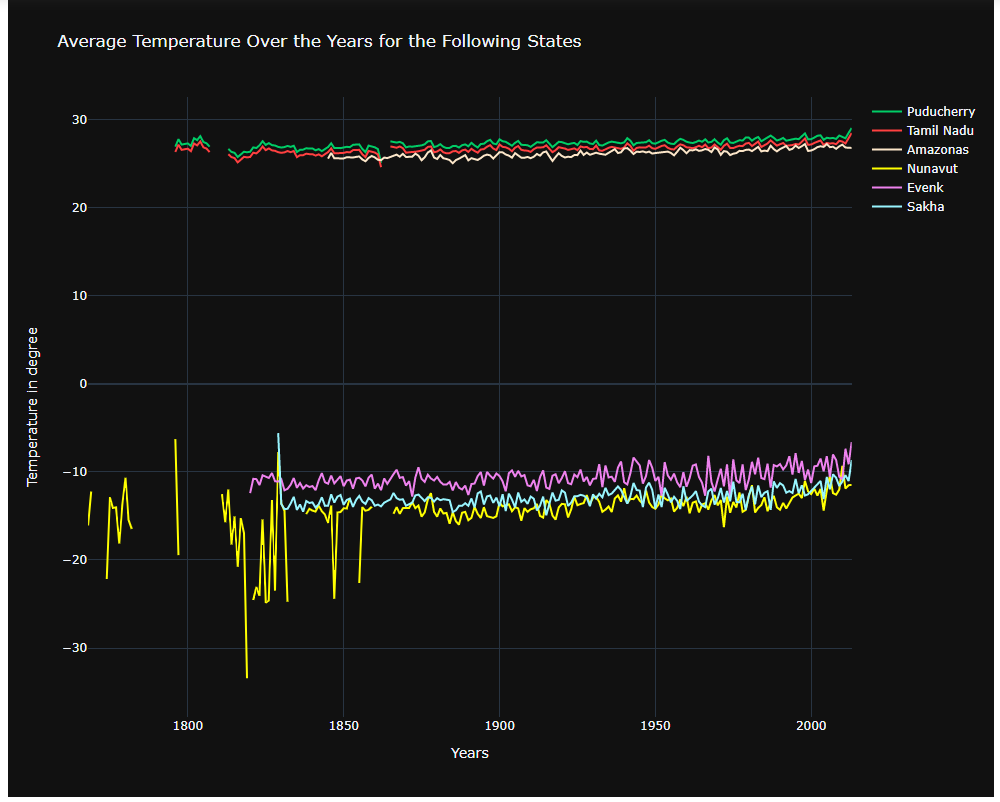
Средняя температура данных 7 стран:





Теперь рассмотреим некоторые из самых горячих и холодных состояний, чтобы увидеть изменение температуры.Рассмотрим Пудучерри, Тамилнад из Индии, Амазонас из Бразилии, Нунавут из Канады, эвенкию и Саху из России.





Видно, что:

Пудучерри имеет разницу в 1,88 градуса за эти годы.

Тамилнад имеет разницу в 1,56 градуса за эти годы.

Амазонас имеет разницу в 1,28 градуса за эти годы.

Нунавут имеет разницу в 4,6 градуса за эти годы.

Эвенкия имеет разницу в 5,8 градуса за эти годы.

Саха имеет разницу в 3,05 градуса за эти годы.

Можем сказать, что в холодных странах температура поднимается быстрее, чем в харких, который считаестя признаком глобального потепления.

**Список литературы**

[1] Гапанюк Ю. Е. Лабораторная работа «Разведочный анализ данных. Исследование  
и визуализация данных» [Электронный ресурс] // GitHub. –– 2019. –– Режим доступа: https://github.com/ugapanyuk/ml\_course/wiki/LAB\_EDA\_VISUALIZATION (дата  
обращения: 13.02.2019)

[2] <https://www.kaggle.com/datasets>