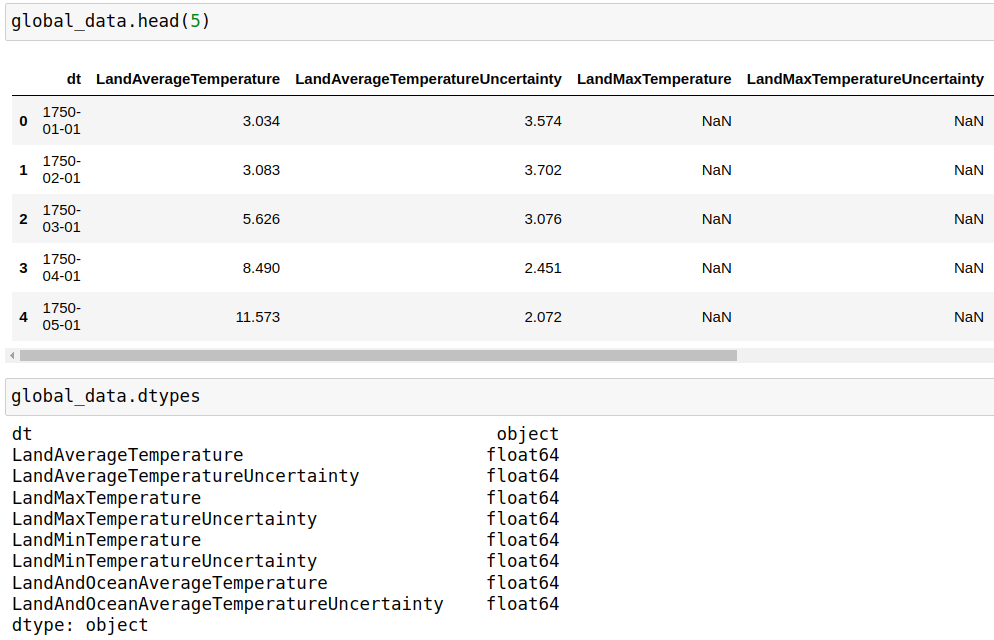
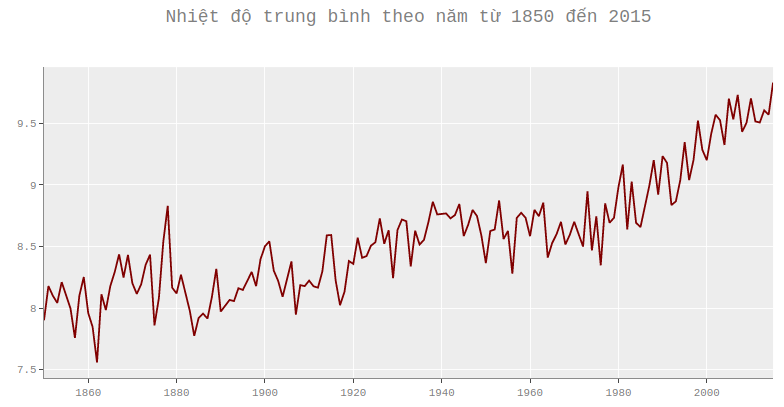
1. **Lời mở đầu**
2. **Sự thay đổi nhiệt độ trên quy mô toàn cầu:**
3. ***Nguồn dữ liệu và xử lý dữ liệu:*** 
   * Để phân tích nhiệt độ trên quy mô toàn cầu, chúng em sử dụng bộ Dữ liệu nhiệt độ bề mặt Trái Đất *Global Temperatures* của [Berkeley Earth](https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data)



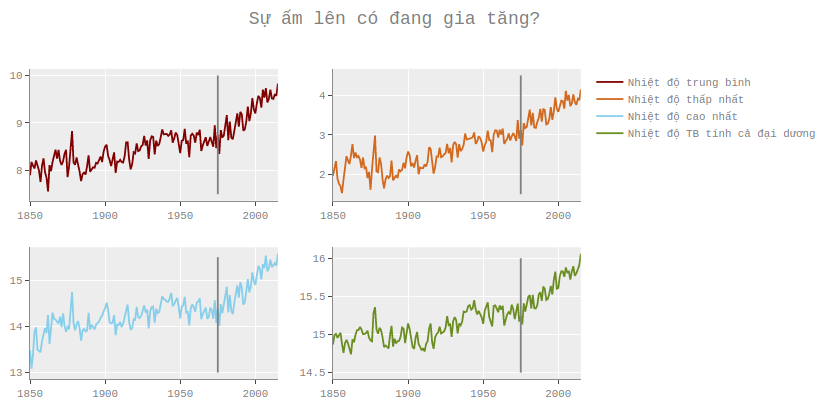
* Vì dữ liệu bị thiếu nằm liên tiếp nhau và đang xử lý với dữ liệu dạng chuỗi thời gian, nên cách hợp lý là xóa toàn bộ các điểm dữ liệu mà có trường thông tin quan trọng bị thiếu khỏi dữ liệu phân tích.
* Xử lý trường thời gian trong dữ liệu: Vì các điểm dữ liệu được thống kê theo ngày trong tổng cộng 166 năm tính từ 1850 và rất nhiều điểm dữ liệu bị thiếu, nên nhóm các điểm dữ liệu về theo năm. Như vậy các trường dữ liệu nhiệt độ sẽ chuyển về nhiệt độ trung bình theo từng năm.
* Để thuận tiện cho phân tích, chúng em tạo thêm trường dữ liệu thời gian tính theo năm (*Year*) và loại bỏ trường dữ liệu *dt* không còn cần thiết.

1. ***Trái Đất có đang ấm dần lên hay không?***
   * Tiến hành lập biểu đồ Line Chart để xem xét xu hướng thay đổi của nhiệt độ bề mặt Trái Đất. Trục x là năm (*Year*), trục y là nhiệt độ trung bình năm (*LandAverageTemperature*).

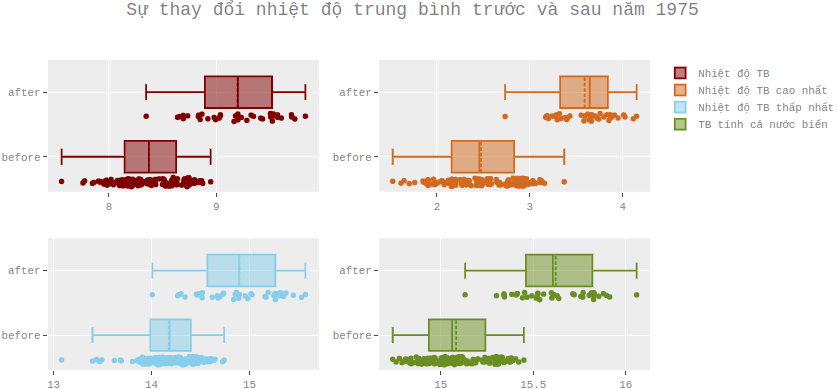


* + Từ biểu đồ dữ liệu nhiệt độ trung bình hàng năm của bề mặt Trái Đất, ta có thể thấy rõ xu hướng tăng của nhiệt độ bề mặt Trái Đất. Nhiệt độ Trái Đất dao động quanh khoảng 8 độ ở nửa cuối thế kỷ 19 và đầu thế kỷ 20. Sau đó nhiệt độ tăng dần qua mốc 9 độ và ở mức trung bình gần 10 độ vào cuối những năm 2000.

1. ***Sự ấm lên toàn cầu có đang gia tăng hay không?***

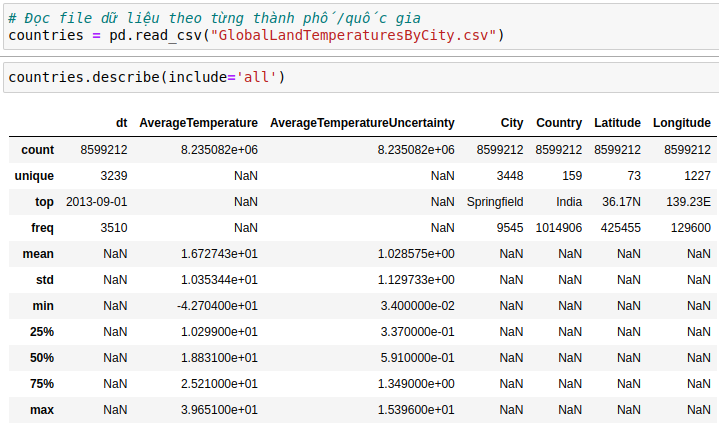


* + Từ 4 biểu đồ nhiệt độ trung bình, lớn nhất, nhỏ nhất cũng như kết hợp cả nhiệt độ nước biển, ta có thể nhận thấy một sự gia tăng nhanh chóng về mức tăng nhiệt độ xảy ra vào khoảng những năm 1970. Đây trùng hợp thay chính là giai đoạn xuất hiện cuộc cách mạng khoa học kỹ thuật và toàn cầu hóa trên thế giới. Cùng với sự phát triển nhanh chóng, mức phát thải khí nhà kính cũng bùng nổ dẫn tới sự gia tăng nhanh hơn của nhiệt độ toàn cầu.

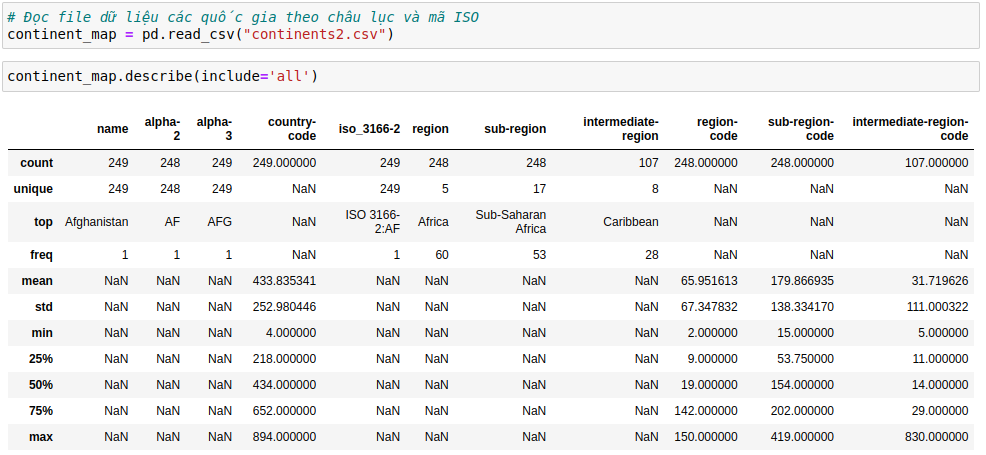


* + Biểu đồ này sẽ thể hiện rõ hơn về sự gia tăng nhiệt độ xảy ra sau thập kỷ 1970. Mức chênh lệch về sự tăng nhiệt trước và sau giai đoạn này lên tới hơn 1 độ.

1. **Sự thay đổi nhiệt độ theo từng khu vực địa lý:**
2. Nguồn dữ liệu:
   * Để phân tích nhiệt độ theo từng khu vực địa lý, chúng em sử dụng hai bộ dữ liệu khác nhau:
   * Bộ thứ nhất là Dữ liệu nhiệt độ bề mặt Trái Đất theo thành phố *GlobalLandTemperaturesByCity.csv* của [Berkeley Earth](https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data)



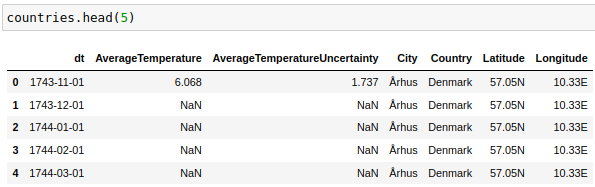
* Bộ dữ liệu này gồm có hơn 8.5 triệu điểm dữ liệu phân bố trên 3448 thành phố khác nhau trên thế giới. Dữ liệu được thống kê theo ngày, bắt đầu từ 1/11 năm 1743 và kết thúc vào 1/9 năm 2013.
* Bộ thứ hai là Dữ liệu các quốc gia theo châu lục *continents2.csv* của[Country Mapping – ISO, Continent, Region](https://www.kaggle.com/datasets/andradaolteanu/country-mapping-iso-continent-region)

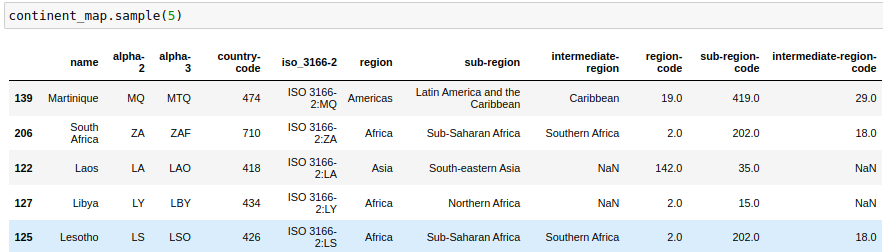


* Bộ dữ liệu này gồm có 249 điểm dữ liệu tương ứng với 249 quốc gia phân bố trên 5 châu lục.

1. Các bước tiền xử lý dữ liệu:

* Đọc dữ liệu từ file:

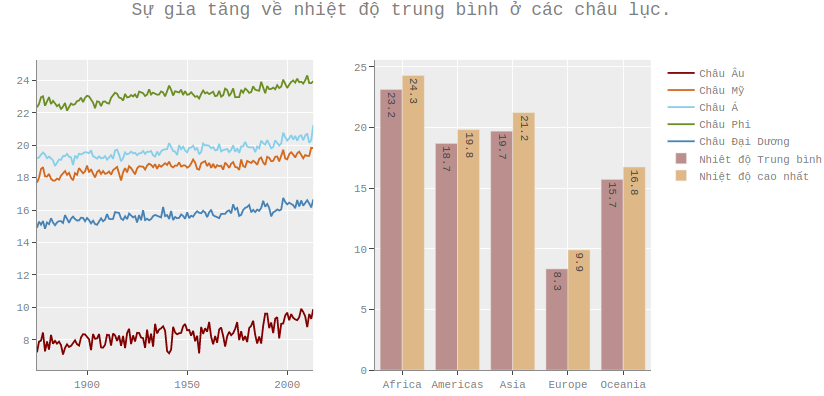




* Loại bỏ các dữ liệu bị thiếu: Vì dữ liệu bị thiếu nằm liên tiếp nhau và đang xử lý với dữ liệu dạng chuỗi thời gian, nên cách hợp lý là xóa toàn bộ các điểm dữ liệu mà có trường thông tin quan trọng bị thiếu.
* Xử lý trường thời gian trong dữ liệu: Vì các điểm dữ liệu được thống kê theo ngày trong tổng cộng 166 năm tính từ 1850 và rất nhiều điểm dữ liệu bị thiếu, nên nhóm các điểm dữ liệu về theo năm.
* Gộp 2 bảng dữ liệu: Gộp hai bảng dữ liệu để thuận tiện cho việc thao tác, phân tích và trực quan hóa.
* Loại bỏ giai đoạn dữ liệu bị thiếu nhiều để giảm sai số khi phân tích thống kê: Loại bỏ các dữ liệu có trước năm 1875 vì trước thời điểm này, nhiều quốc gia không có dữ liệu đo đạc nên sẽ gây ảnh hưởng đến kết quả phân tích.

1. Mức thay đổi nhiệt độ có đồng nhất giữa các châu lục hay không?

Chúng em tiến hành trực quan hóa dữ liệu nhiệt độ trung bình trên 5 châu lục theo năm để quan sát sự biến thiên về nhiệt độ giữa các châu lục này.

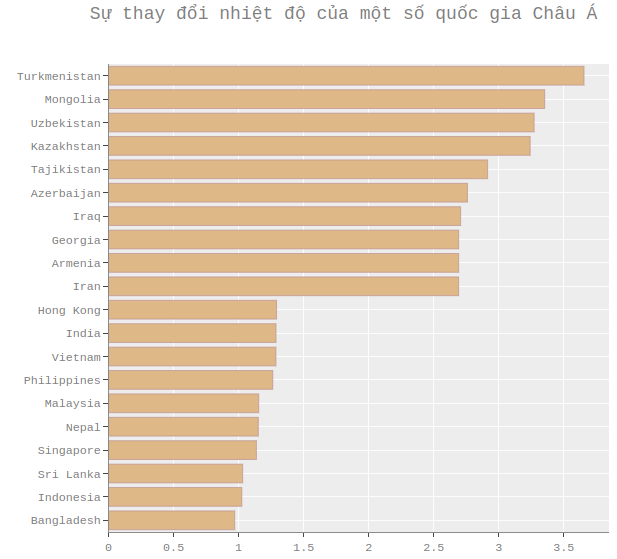


Biểu đồ cho thấy nhiệt độ trung bình trên khắp các châu lục có xu hướng tăng dần theo từng năm. Tuy nhiên mức tăng nhiệt độ là không đồng đều giữa các châu lục.

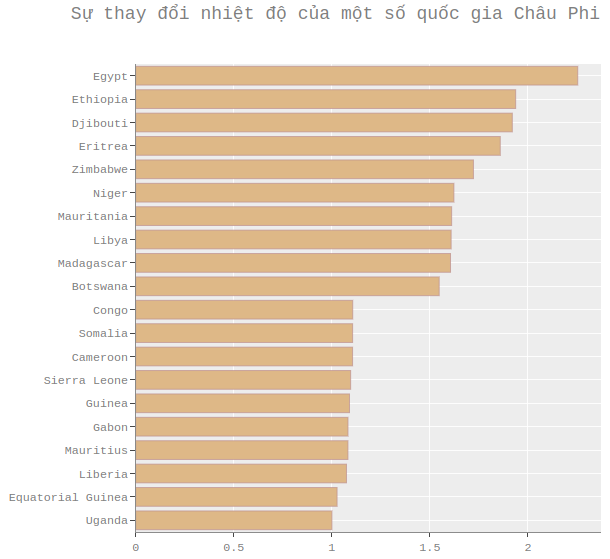
* Châu Âu mặc dù có nền nhiệt độ trung bình thấp nhất trong số các châu lục, chỉ 8.3 độ, tuy nhiên lại chứng kiến mức gia tăng nhiệt độ lớn nhất, với chênh lệch 1.6 độ giữa mức cao nhất và mức trung bình.
* Châu Phi mặc dù có nhiệt độ trung bình cao nhất trong 5 châu lục, ở mức 23.2 độ, nhưng mức tăng nhiệt chỉ vào khoảng 1.1 độ, bằng với mức tăng nhiệt của Châu Đại Đương và Châu Mỹ.
* Ở biểu đồ đường, ta thấy Châu Đại Dương có nhiệt độ biến động mạnh hơn và có vẻ mang tính chu kỳ, đây có thể là do hiệu ứng của hiện tương La Nina.
* Nhìn chung, tất cả các châu lục đều chứng kiến nhiệt độ gia tăng với mức dao động từ 1 cho đến 1.5 độ.

1. Vị trí địa lý của các quốc gia ở mỗi châu lục ảnh hưởng thế nào đến mức thay đổi nhiệt độ?

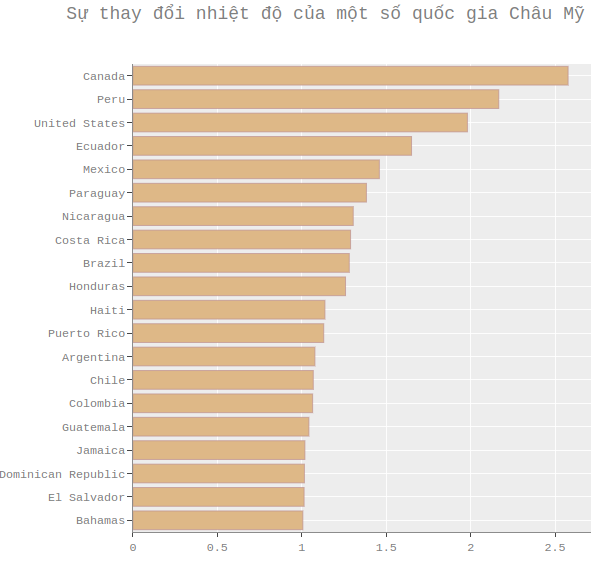
* Chúng em tiến hành khảo sát 10 quốc gia có mức tăng nhiệt lớn nhất và nhỏ nhất ở từng châu lục để quan sát ảnh hưởng của vị trí địa lý đến nhiệt độ. Riêng Châu Đại Dương không xét vì chỉ có 3 quốc gia được khảo sát.



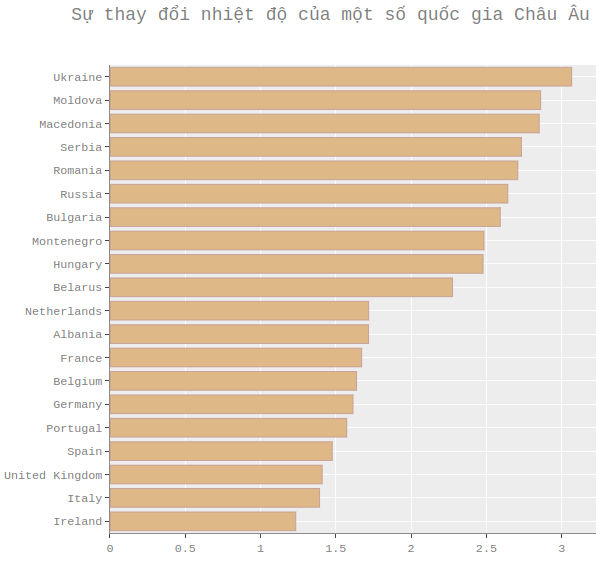
* Ở Châu Á, các quốc gia chịu ảnh hưởng lớn nhất của sự nóng lên toàn cầu là các quốc gia nằm ở khu vực Trung và Tây Á với mức tăng lên tới 3.5 độ ở Turkmenistan. Trong khi đó các quốc gia ở khu vực Nam Á và Đông Nam Á có mức tăng nhiệt thấp hơn chỉ khoảng 1 độ.

****

* Tương tự như ở Châu Á, các quốc gia Nam Phi có mức tưng nhiệt thấp hơn đáng kể các quốc gia ở Bắc và Trung Phi. Ai Cập là quốc gia có mức tăng lớn nhất với hơn 2 độ.

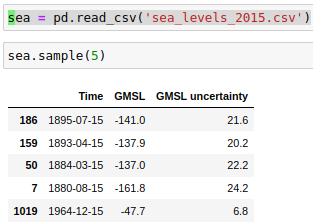


* Ở Châu Mỹ, ta có thể thấy Bắc Mỹ chịu ảnh hưởng lớn nhất với mức tăng nhiệt lên tới hơn 2.5 độ diễn ra ở Canada. Các quốc gia Trung và Nam Mỹ có mức tăng chênh nhau không đáng kể.

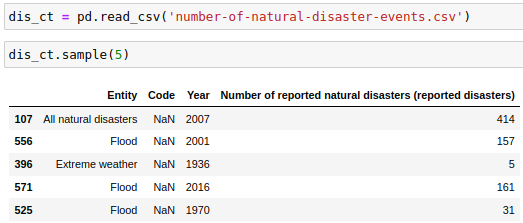


* Tại Châu Âu, các quốc gia Đông Âu chịu ảnh hưởng lớn hơn từ sự nóng lên toàn cầu với Ukraine ảnh hưởng nặng nhất với mức tăng lên tới hơn 3 độ.
* Như vậy, ta có thể thấy rằng, vị trí địa lý là một nhân tố quan trọng ảnh hưởng tới mức tăng nhiệt độ của từng quốc gia, với Trung và Tây Á chịu ảnh hưởng nặng nề nhất và Bán Cầu Bắc chịu ảnh hưởng nhiều hơn Bán Cầu Nam.

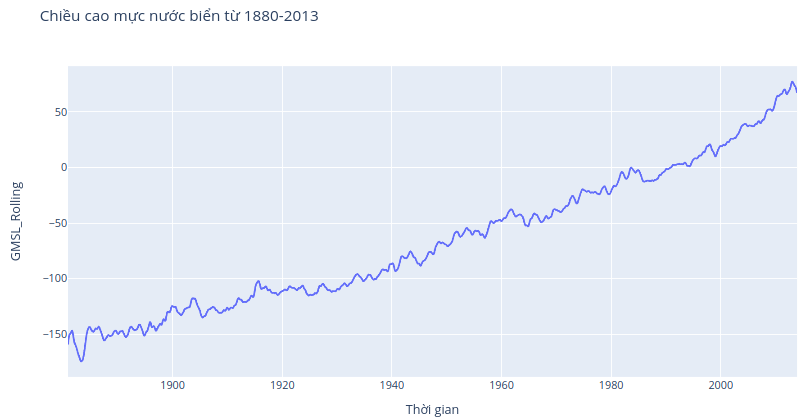
1. **Ảnh hưởng của dự gia tăng nhiệt độ tới các hiện tượng tự nhiên:**
2. Dữ liệu và xử lý dữ liệu:
   * Để xem xét ảnh hưởng của sự ấm lên toàn cầu với các hiện tượng tự nhiên khác như mực nước biển hay các thảm họa tự nhiên, chúng em sử dụng thêm hai bộ dữ liệu khác là dữ liệu [Sea Level Change](https://www.kaggle.com/datasets/somesh24/sea-level-change) của Cơ quan bảo vệ môi trường Mỹ và dữ liệu [Natural Disaster](https://www.kaggle.com/datasets/dataenergy/natural-disaster-data) của tổ chức [Our World in Data](https://ourworldindata.org/natural-disasters).
   * Do dữ liệu mực nước biển được thông kê theo tháng nên sẽ bị ảnh hưởng bởi các yếu tố mang tính chu kỳ trong năm. Vì vậy với mực nước biển, chúng em sử dụng phương pháp trung bình cho rolling 12 tháng.



* + Ở dữ liệu thảm họa tự nhiên, chúng em chỉ sử dụng thông tin về số lượng của các thảm họa và năm diễn ra, nên sẽ loại bỏ đi các cột còn lại.

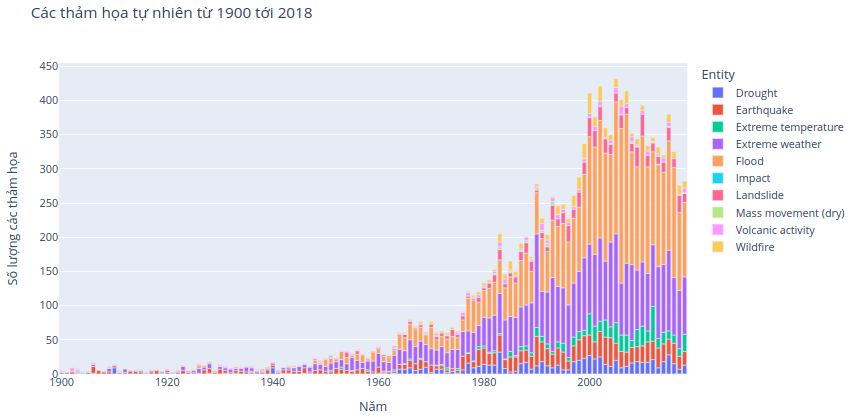


1. Quan hệ giữa sự gia tăng nhiệt độ và mực nước biển

****

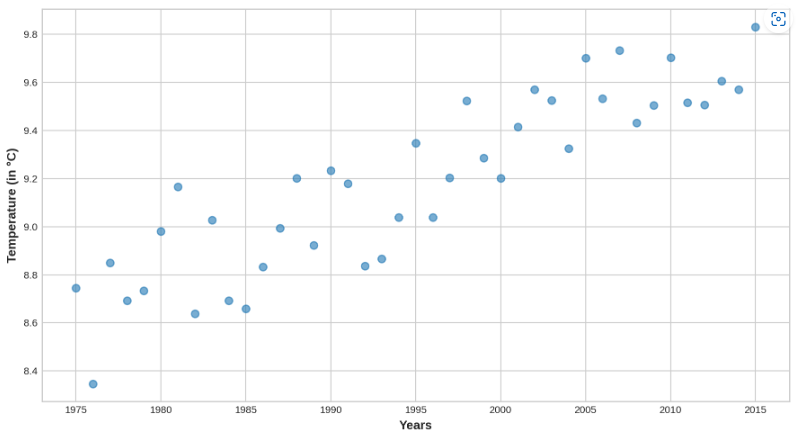
* + Từ biểu đồ mực nước biển theo năm bên trên, ta có thể thấy rõ một xu hướng gia tăng về chiều cao mực nước biển, xu hướng này là đồng biến với xu hướng ấm lên của Trái Đất. Nguyên nhân có thể là do nhiệt độ ấm lên làm gia tăng sự tan băng ở hai cực của Trái Đất, giải phóng lượng nước khổng lồ ra các đại dương.

1. Sự gia tăng nhiệt độ và các tần suất các thảm họa tự nhiên

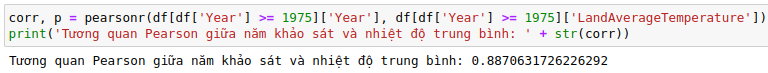
****

* + Rõ ràng có thể thấy từ biểu đồ, các thảm họa tự nhiên đã gia tăng dần theo từng năm và bứt tốc nhanh chóng vào khoảng thập niên 1970. Trùng hợp là kể từ thập niên 1970, nhiệt độ Trái Đất cũng có sự gia tăng đáng kể so với giai đoạn trước đó. Điều này có thể phần nào phản ánh sự tương quan giữa sự ấm lên của Trái Đất và sự gia tăng tần suất của các thảm họa tự nhiên.

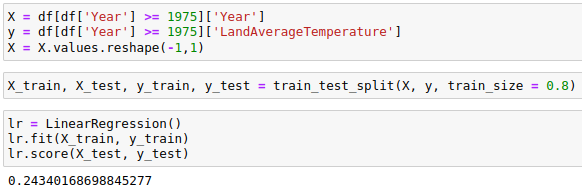
1. **Dự đoán xu hướng thay đổi của nhiệt độ:**
2. Xử lý dữ liệu: Do sau năm 1975 có một sự thay đổi đột ngột về xu hướng tăng nhiệt độ toàn cầu, nên chúng em chỉ xét các điểm dữ liệu tính từ 1975 để làm căn cứ dự đoán nhiệt độ tương lai.
3. Xem xét xu hướng và dự đoán:

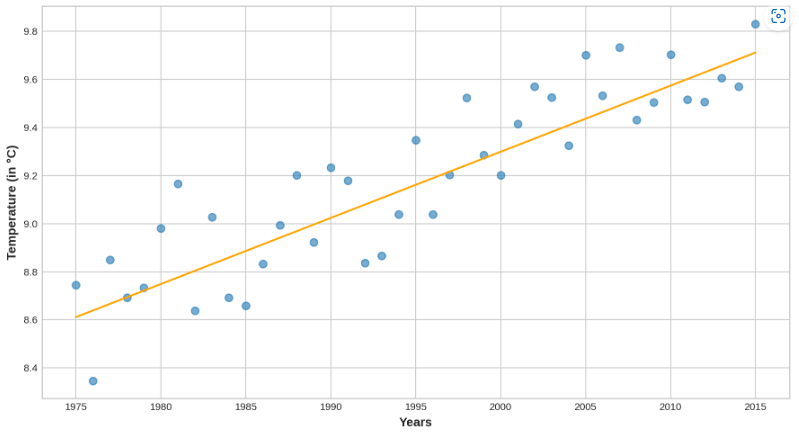


* + Trực quan hóa mối quan hệ giữa hai thuộc tính nhiệt độ trung bình và năm cho thấy có xu hướng tương quan tuyến tính giữa hai biến này.



* + Tính toán hệ số tương quan Pearson giữa hai biến cho thấy chúng có một mối tương quan mạnh.
  + Chúng em fit một mô hình hồi quy tuyến tính đơn cho dữ liệu để dự đoán nhiệt độ của Trái Đất trong tương lai



****

* Tiến hành dự đoán nhiệt độ của các năm 2030, 2040 và 2050:

****

* Nếu xu hướng tăng nhiệt độ vẫn tiếp diễn như hiện nay, thì nhiệt độ trung bình của Trái Đất sẽ là 10.1 độ vào năm 2030 và lên tới 10.7 độ vào giữa thế kỷ này.

1. **Kết luận:**