

Russian Federation: A Discriminant Analysis on the Paleoclimatology in Russia

Julia Cuellar

Summer 2021

<https://github.com/BVUjac8/Russian-Federation.git>

Which Domain?

Data.gov

10 Key Types of Data Analysis Methods and Techniques. (n.d.). Retrieved from [10 Top Types of Data Analysis Methods and Techniques \(intellspot.com\)](#)

Lake Baikal. (n.d.). Retrieved from [Lake Baikal - Wikipedia](#)

Lake Baikal: The Pearl of Siberia. (n.d.). Retrieved from [Lake Baikal in Siberia, Russia | Baikal Lake Facts & Map \(lakepedia.com\)](#)

Lake Elgygytyn. (n.d.). Retrieved from [Lake Elgygytyn - Wikipedia](#)

NOAA/WDS Paleoclimatology – Lake Baikal Composite BDP-96 Pleistocene Biogenic Silica Data. (n.d.). Retrieved from [NOAA/WDS Paleoclimatology - Lake Baikal Composite BDP-96 Pleistocene Biogenic Silica Data - CKAN](#)

NOAA/WDS Paleoclimatology – Lake El'gygytyn, NE Russia Quaternary Multiproxy Lake Sediment Data. (n.d.). Retrieved from [NOAA/WDS Paleoclimatology - Lake El'gygytyn, NE Russia Quaternary Multiproxy Lake Sediment Data - CKAN](#)

Pleistocene. (n.d.). Retrieved from [Pleistocene - Wikipedia](#)

Pleistocene Epoch. (n.d.). Retrieved from [Pleistocene Epoch | Characteristics, Plants, Animals, Climate, & Facts | Britannica](#)

Strauss, B. (2020, January 21). *Prehistoric Life During the Pleistocene Epoch*. [Prehistoric Life During the Pleistocene Epoch \(thoughtco.com\)](#)

The Thrill to Drill in the Chill – 3.6 Million Years of Arctic Climate Change from Lake El'gygytyn, NE Russia. (n.d.). Retrieved from [Lake El'gygytyn \(dosecc.com\)](#)

Vermillion, S. (2021, April 16). *Siberia's Lake Baikal Is the World's Oldest and Weirdest*. [Siberia's Lake Baikal Is the World's Oldest and Weirdest | HowStuffWorks](#)

Which Data?

<https://www.ncei.noaa.gov/pub/data/paleo/paleolimnology/asia/russia/elgygytyn2012.txt>

Data from Data.gov from the National Oceanic and Atmospheric Administration & the Department of Commerce.

A dataframe with 91 observations and 9 variables.

**Age..ka. MTWM..deg.C. MTWM...deg.C. MTWM...deg.C..1 PANN..mm. PANN...mm.
PANN...mm..1 Trees...Shrubs.... Picea**

A dataframe with 99 observations and 19 variables.

**Age..ka.BP. Picea.sect..Eupicea P..s.g.Haplo.T Larix Betula Alnus Salix Poaceae Cyperaceae
Artemisia Ericales Cphae**

<https://www.ncei.noaa.gov/pub/data/paleo/paleolimnology/asia/russia/baikal2006.txt>

Data from Data.gov from the National Oceanic and Atmospheric Administration & the Department of Commerce.

A dataframe with 3,669 observations and 6 variables.

Hole.....Core.Sec..Int.....Depth.....Bio.sil.....Age

Research Questions? Benefits? Why analyze these data?

- What is the common column variable that can be made from these 3 different file formats (txt, JSON, & xls) to merge into a database?

Having worked with CSV files, a Web page, and an API in Python, the same approach (now in R programming) to merge three different file formats into a single dataset and transferring it into a database based off a common column variable should be implemented to understand various file formats and working with them in relation to a database interface.

What Method?

There is only one method to employ when just working with 3 different file formats which is in the form of discriminant analysis. Discriminant analysis is a classification technique that utilizes variable measurements on different groups of items to underline points that distinguish the groups.

Potential Issues?

Cleaning the data will be the most hurdle due to working with not just one but 3 datasets. Although, another setback could be working in R programming with different file formats I am unused to manipulating.

Concluding Remarks

As potential data scientists, we must be accustomed to working with almost all different file formats as well as an alternative programming language. This project will see to that I have broaden my understanding in both R programming as well as wielding txt, JSON, and xls file formats.