

DSCI 310: Historical Horse Population in Canada

Tiffany Timbers & Jordan Bourak

Table of Contents

1	Aim	1
2	Data	1
3	Methods	2
4	Results	2

```
import pandas as pd
from IPython.display import Markdown, display
from tabulate import tabulate
```

1 Aim

This project explores the historical population of horses in Canada between 1906 and 1972 for each province.

2 Data

Horse population data were sourced from the [Government of Canada's Open Data website](#) (Government of Canada, 2017a and Government of Canada, 2017b).

Table 1

Table 1: Standard deviation of historical (1906-1972) horse populations for each Canadian province.

3 Methods

The Python programming language (Van Rossum and Drake 2009) and the following Python packages were used to perform the analysis: pandas (McKinney 2010), altair (VanderPlas 2018), click (Team 2020), as well as Quarto (Allaire et al. 2022).

Note: this report is adapted from Timbers (2020).

4 Results

We can see from Figure Figure 1 that Ontario, Saskatchewan, and Alberta have had the highest horse populations in Canada.

All provinces have had a decline in horse populations since 1940.

This is likely due to the rebound of the Canadian automotive industry after the Great Depression and the Second World War.

An interesting follow-up visualisation would be car sales per year for each Province over the time period visualised above to further support this hypothesis.

Suppose we were interested in looking more closely at the province with the highest spread (in terms of standard deviation) of horse populations.

We present the standard deviations in Table Table 1.

```
horses_sd_table = pd.read_csv("../results/horses_sd.csv")
largest_sd = horses_sd_table['Province'].values[0]
Markdown(horses_sd_table.to_markdown(index=False))
```

Province	Std
Saskatchewan	377266
Ontario	266435
Alberta	266063
Manitoba	122404
Quebec	111411
New Brunswick	22019.5
Nova Scotia	19879.3

Province	Std
British Columbia	14945.7
P.E.I.	11355.7

Note that we define standard deviation (of a sample) as:

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$$

Additionally, note that in **Table Table 1** we consider the sample standard deviation of the number of horses during the same time span as Figure Figure 1.

Figure Figure 2 shows the province with the largest standard deviation.

In Figure Figure 2, we zoom in and look at the province of `r_largest_sd`, which had the largest spread of values in terms of standard deviation.

References ## References

- Allaire, J. J., Charles Teague, Carlos Scheidegger, Yihui Xie, and Christophe Dervieux. 2022. “Quarto.” <https://doi.org/10.5281/zenodo.5960048>.
- McKinney, Wes. 2010. “Data Structures for Statistical Computing in Python.” In *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman, =51–56.
- VanderPlas, Jake. 2018. “Altair: Interactive Statistical Visualizations for Python.” *Journal of Open Source Software* 3 (7825, 32): 1057. <https://doi.org/10.21105/joss.01057>.

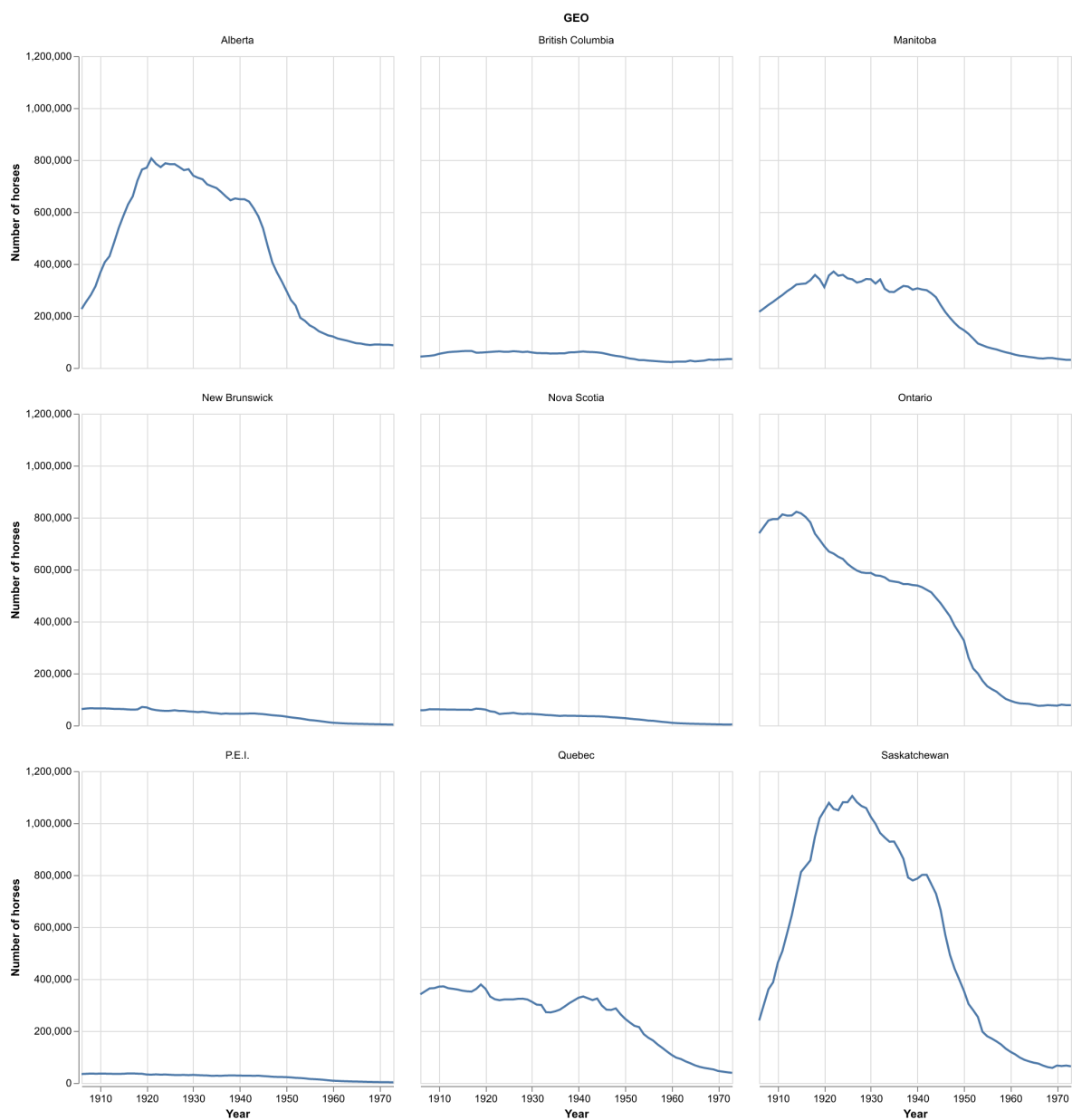


Figure 1: Horse populations for all provinces in Canada from 1906 - 1972

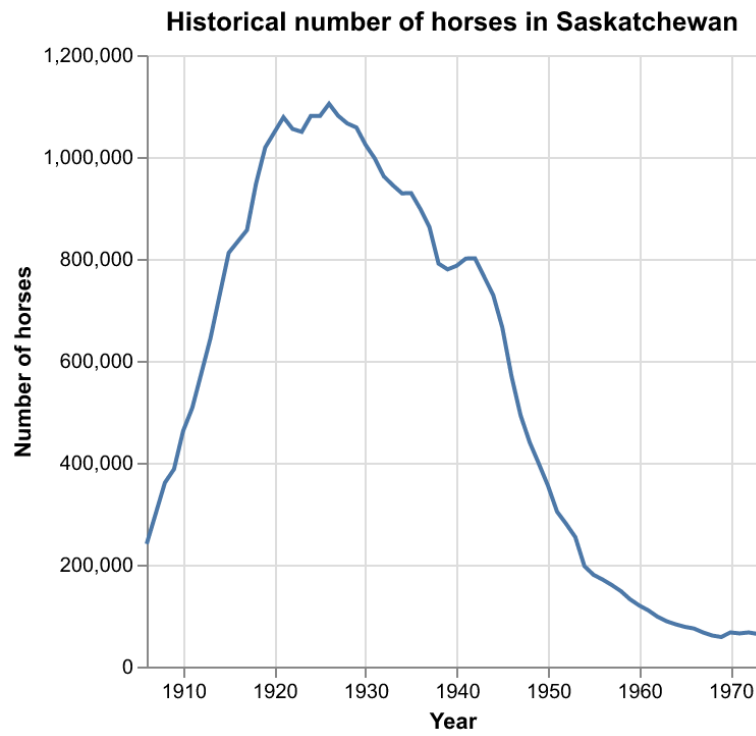


Figure 2: Horse populations for the province with the largest standard deviation