

DSCI 310: Historical Horse Population in Canada

Tiffany Timbers & Jordan Bourak

Table of contents

| | |
|-------------------|----------|
| Aim | 1 |
| Data | 1 |
| Methods | 2 |
| Results | 2 |
| References | 3 |

Aim

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

Data

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

Methods

The analysis was conducted using Python (Van Rossum and Drake 2009) along with several Python packages: pandas for data manipulation (McKinney 2010), Altair for visualization (VanderPlas 2018), and Click for command-line interfaces (Team 2020). This report was rendered with Quarto (Allaire et al. 2022). *Note: This report is adapted from Timbers' earlier work (Timbers 2020).*

Results

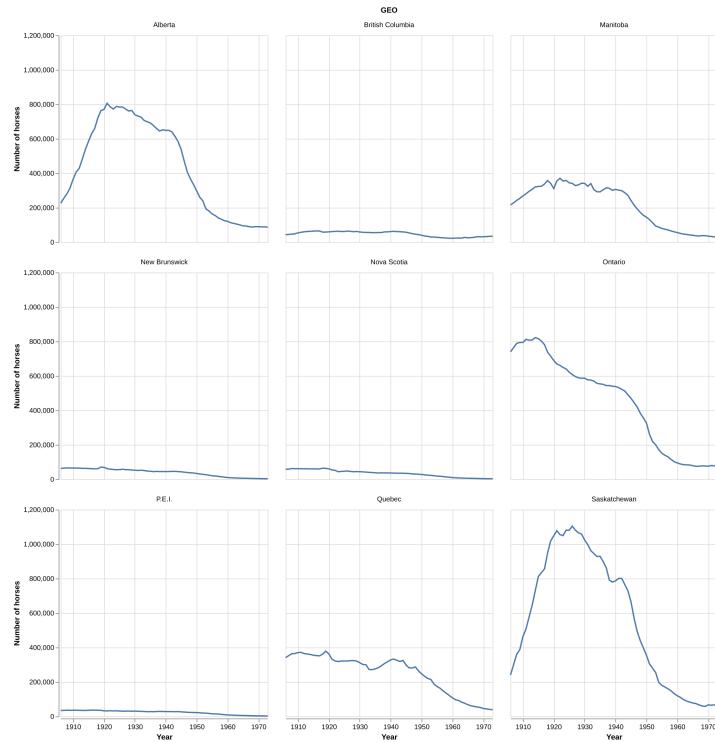


Figure 1: Horse populations for all provinces in Canada from 1906 - 1972.

We can see from 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 in more closely at the province with the highest spread (in terms of standard deviation) of horse populations. We present the standard deviations in

Table 1.

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

| Province | Std |
|------------------|---------|
| Saskatchewan | 377266 |
| Ontario | 266435 |
| Alberta | 266063 |
| Manitoba | 122404 |
| Quebec | 111411 |
| New Brunswick | 22019.5 |
| Nova Scotia | 19879.3 |
| 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 1 we consider the sample standard deviation of the number of horses during the same time span as Figure 1.

From Table 1, we see that the province with the largest standard deviation is Saskatchewan.

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

References

- Allaire, J. J., Charles Teague, Carlos Scheidegger, Yihui Xie, and Christophe Dervieux. 2022. “Quarto.” <https://doi.org/10.5281/zenodo.5960048>.
- Government of Canada. 2017a. “Horses, Number on Farms at June 1 and at December 1.” Open Government - Open Data. <https://open.canada.ca/data/en/dataset/a3ecf553-8ec4-4551-a0fe-8df1472c6cf7>.
- . 2017b. “Horses, Number on Farms at June 1, Farm Value Per Head and Total Farm Value.” Open Government - Open Data. <https://open.canada.ca/data/en/dataset/e175ef9c-98f0-49b3-8131-ca0e3895a0cb>.

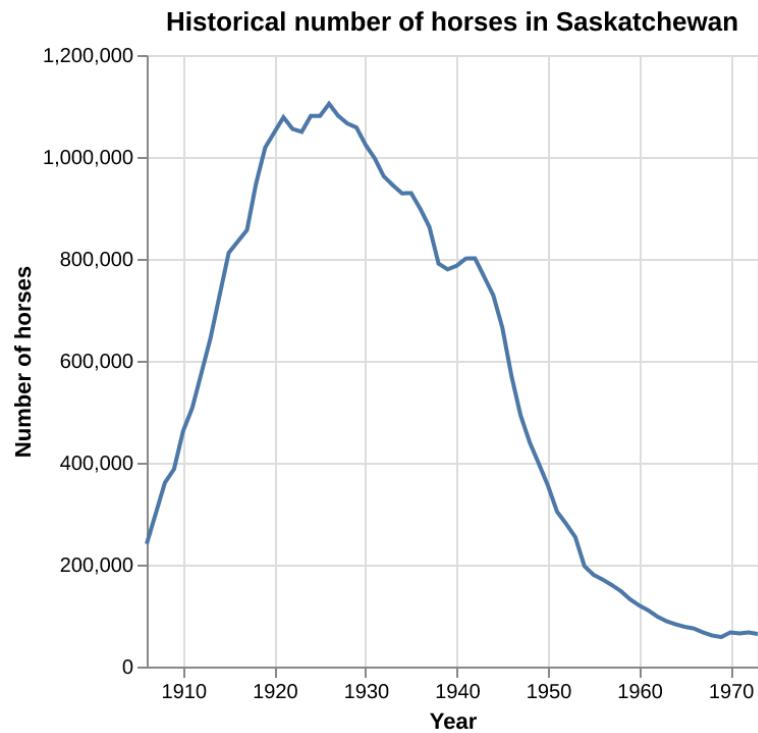


Figure 2: Horse populations for the province with the largest standard deviation.

- 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.
- Team, Pallets. 2020. *Click*. <https://click.palletsprojects.com/>.
- Timbers, Tiffany. 2020. *Historical Horse Population in Canada*. https://github.com/ttimbers/equine_numbers_value_canada_parameters.
- Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- 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>.