

DSCI 310: Historical Horse Population in Canada

Nika Karimi Seffat

Table of contents

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

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.2

-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

1 Aim

The aim of this report is to analyze historical trends in the horse population across Canada. By examining open government datasets and applying reproducible data analysis techniques, we seek to identify key demographic shifts and inform future agricultural policies [at timbers-horses].

2 Data

Horse population data were sourced from the [Government of Canada's Open Data website](#) [@horses1; @horses2]

3 Methods

The R programming language [@R] and the following R packages were used to perform the analysis: knitr [@knitr], tidyverse [@tidyverse], and Quarto [@Allaire_Quarto_2022]. *Note: this report is adapted from [@timbers-horses].*

4 Results

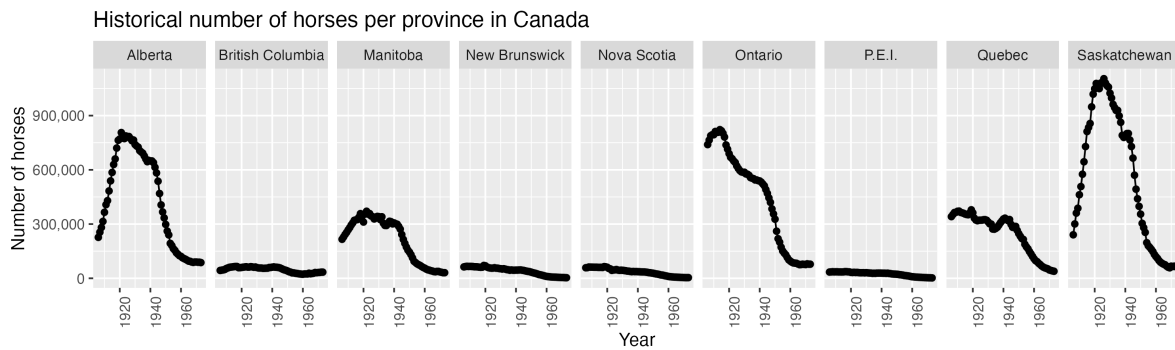


Figure 1: Horse populations for all provinces in Canada from 1906 - 1972.

We can see from @figure-horse 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 @std-dev-table-horses.

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

```

Rows: 9 Columns: 2
-- Column specification -----
Delimiter: ","
chr (1): Province
dbl (1): Std

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

```

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

Province	Std
Saskatchewan	377265.58
Ontario	266435.32
Alberta	266063.19
Manitoba	122403.87
Quebec	111411.10
New Brunswick	22019.49
Nova Scotia	19879.25
British Columbia	14945.66
P.E.I.	11355.75

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 @std-dev-table-horses we consider the sample standard deviation of the number of horses during the same time span as @figure-horse.

In @horse-population-largest-std-dev we zoom in and look at the province of Saskatchewan which had the largest spread of values in terms of standard deviation.

5 References

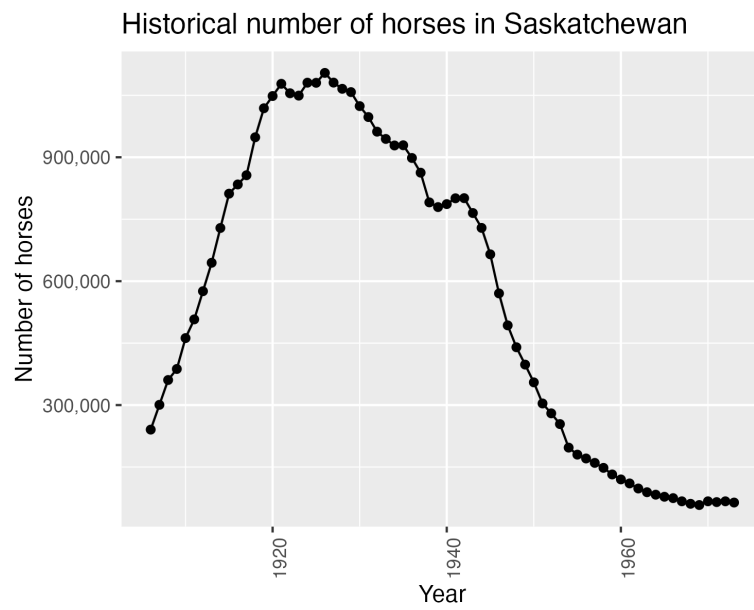


Figure 2: For the province with the biggest standard deviation - the horse population