DSCI 310: Historical Horse Population in Canada

Tiffany Timbers & Jordan Bourak

Mar 16, 2023

CONTENTS

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. Specifically, [the Government of Canada, 2017] 1 and [the Government of Canada, 2017] 2.

Methods

The R programming language [R Core Team, 2019] and the following R packages were used to perform the analysis: knitr [Xie, 2014], tidyverse [Wickham, 2017], and bookdown [Xie, 2016]. *Note: this report is adapted from [Timbers, 2020].*
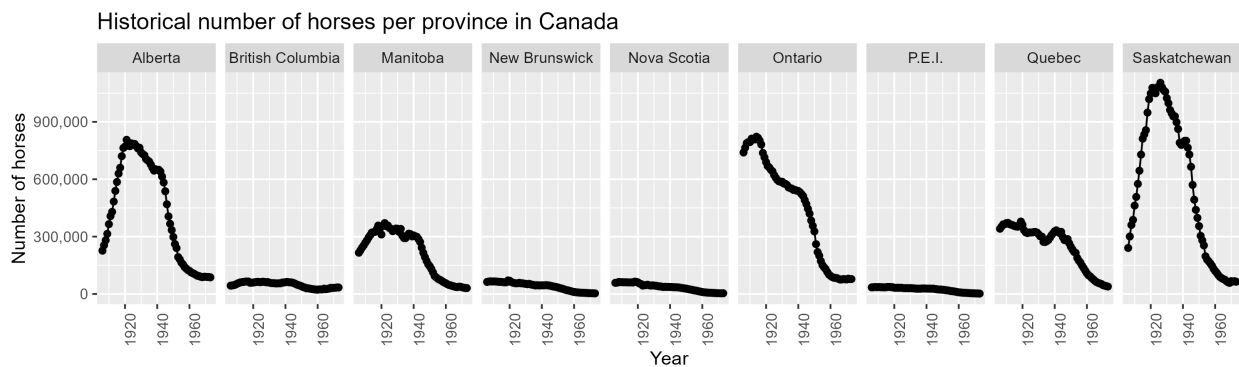
Results

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

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

```
horses_sd = pd.read_csv("../../results/horses_sd.csv")

largest_sd_prov = str(horses_sd['Province'][0])
glue("largest-sd-prov", largest_sd_prov)

horses_sd_noindex = horses_sd.style.hide_index()
glue("horses-tbl", horses_sd_noindex)
```

```
'Saskatchewan'
```

```
C:\Users\Julia Liao\AppData\Local\Temp\ipykernel_9844\765628425.py:6:␣
 ↪FutureWarning: this method is deprecated in favour of `Styler.hide(axis="index")`
   horses_sd_noindex = horses_sd.style.hide_index()
```

```
<pandas.io.formats.style.Styler at 0x172eb4cb100>
```

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 here:

```
<pandas.io.formats.style.Styler at 0x172eb4cb100>
```

Table 1: Standard deviation of number of horses for each province between 1940 - 1972

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

$$s = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})/n - 1}.$$

Additionally, note that in Fig 2. we consider the sample standard deviation of the number of horses during the same time span as Fig. 1.

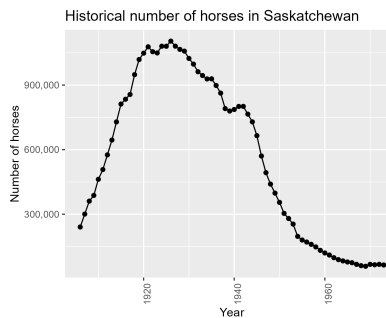

Fig. 2: Horse populations for the province with the largest standard deviation

In Figure 2: Horse populations for the province with the largest standard deviation we zoom in on the province of ?, which had the largest spread of values in terms of standard deviation.

BIBLIOGRAPHY

[tGoC17a]   the Government of Canada. Horses, number on farms at june 1 and at december 1. Open Government - Open Data, 2017. URL: https://open.canada.ca/data/en/dataset/a3ecf553-8ec4-4551-a0fe-8df1472c6cf7.

[tGoC17b]   the Government of Canada. Horses, number on farms at june 1, farm value per head and total farm value. Open Government - Open Data, 2017. URL: https://open.canada.ca/data/en/dataset/e175ef9c-98f0-49b3-8131-ca0e3895a0cb.

[Tim20]   Tiffany Timbers. Historical horse population in canada. 2020. URL: https://github.com/ttimbers/equine_numbers_value_canada_parameters.

[Wic17]   Hadley Wickham. *tidyverse: Easily Install and Load the 'Tidyverse'*. 2017. R package version 1.2.1. URL: https://CRAN.R-project.org/package=tidyverse.

[Xie14]   Yihui Xie. *knitr: A Comprehensive Tool for Reproducible Research in R*. Chapman and Hall/CRC, 2014. ISBN 978-1466561595. URL: http://www.crcpress.com/product/isbn/9781466561595.

[Xie16]   Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman and Hall/CRC, Boca Raton, Florida, 2016. ISBN 978-1138700109. URL: https://bookdown.org/yihui/bookdown.

[RCTeam19] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL: https://www.R-project.org/.