
DSCI 310: Historical Horse Population in Canada

Tiffany Timbers & Jordan Bourak

Mar 18, 2022

CONTENTS

Bibliography	3
---------------------	----------

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, [[Government of Canada, 2017](#), [Government of Canada, 2017](#)]

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

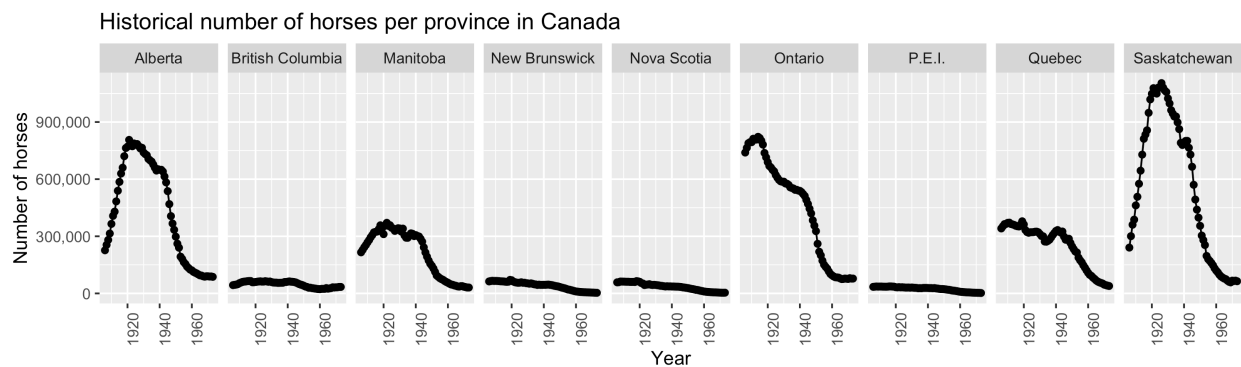


Fig. 1: Horse populations for all provinces in Canada from 1906 - 1972

We can see from [Figure 1](#) 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, display=False)

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

```
/var/folders/h0/smwkvjnx5njf00v9_bs7wjvc0000gn/T/ipykernel_75291/1568206956.py:6:
FutureWarning: this method is deprecated in favour of `Styler.hide(axis='index')`
horses_sd_noindex = horses_sd.style.hide_index()
```

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 0x1037901f0>
```

Fig. 2: Standard deviation of number of horses for each province between 1940 - 1972

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

$$s = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1)}.$$

Additionally, note that in [Figure 2](#) we consider the sample standard deviation of the number of horses during the same time span as [Figure 1](#).

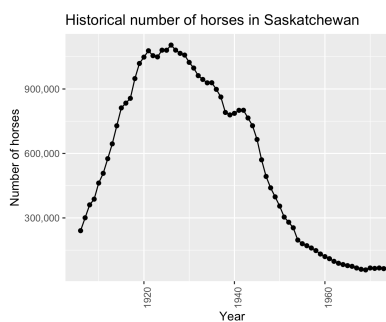


Fig. 3: Horse populations for the province with the largest standard deviation

In [Figure 3](#) we zoom in on the province of Saskatchewan, which had the largest spread of values in terms of standard deviation.

References

BIBLIOGRAPHY

- [Wic17] **missing publisher in 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>.
- [GovernmentoCanada17a] **missing journal in horses1**
- [GovernmentoCanada17b] **missing journal in horses2**
- [RCTeam19] **missing publisher in R**