

R Course: lesson 2

Cédric HASSEN-KHODJA

14 juin 2016

Introduction

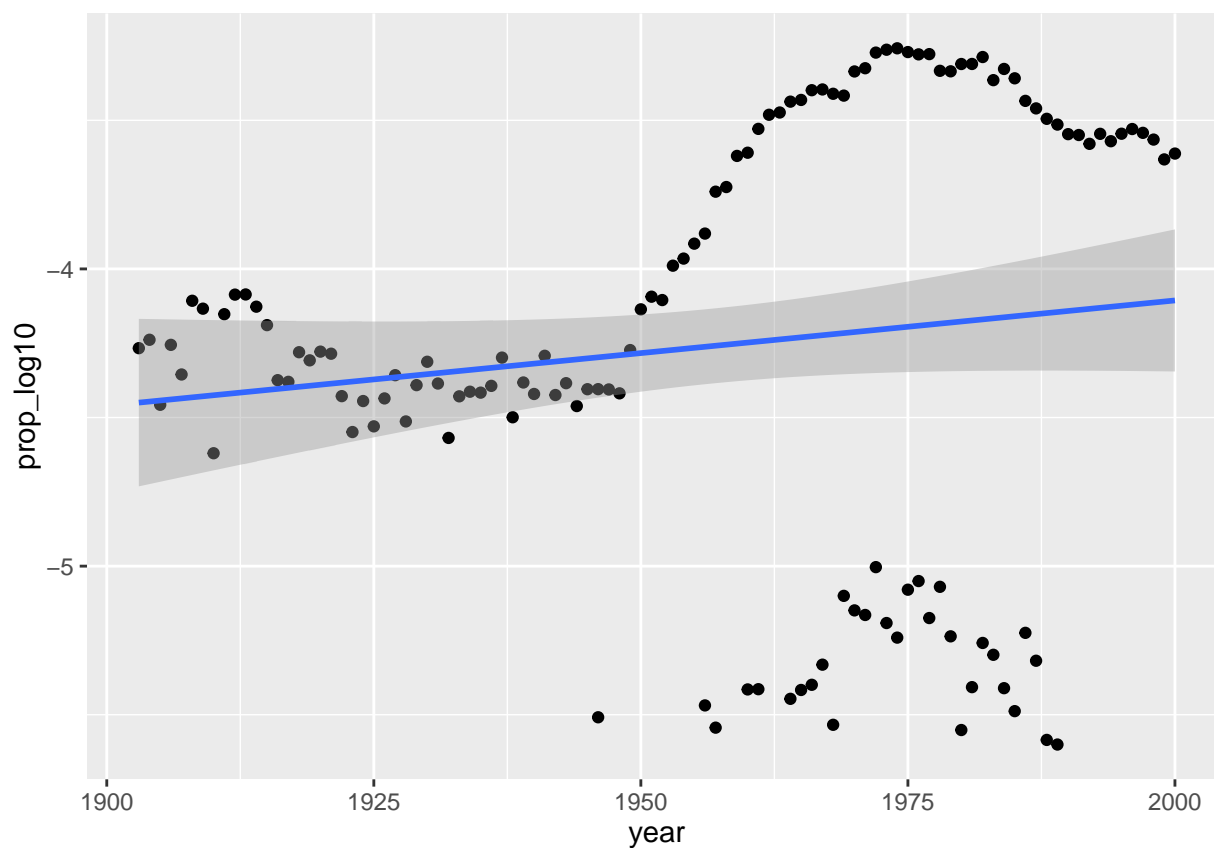
I looked at how common my name, “Cedric”, is in the United States population both by year and sex.

Results

Prevalence by Year

Below is a plot for how the proportion of people with the name “Cedric” (log base 10 transformed) has changed over time. Data aren’t normal in this case. Interpretation of result is difficult.

```
year.plot
```



To test if there is a significant effect of year a linear model was built. Proportion of the population log base 10 transformed was the dependent variable and year the independent variable. As shown below, year was not significant.

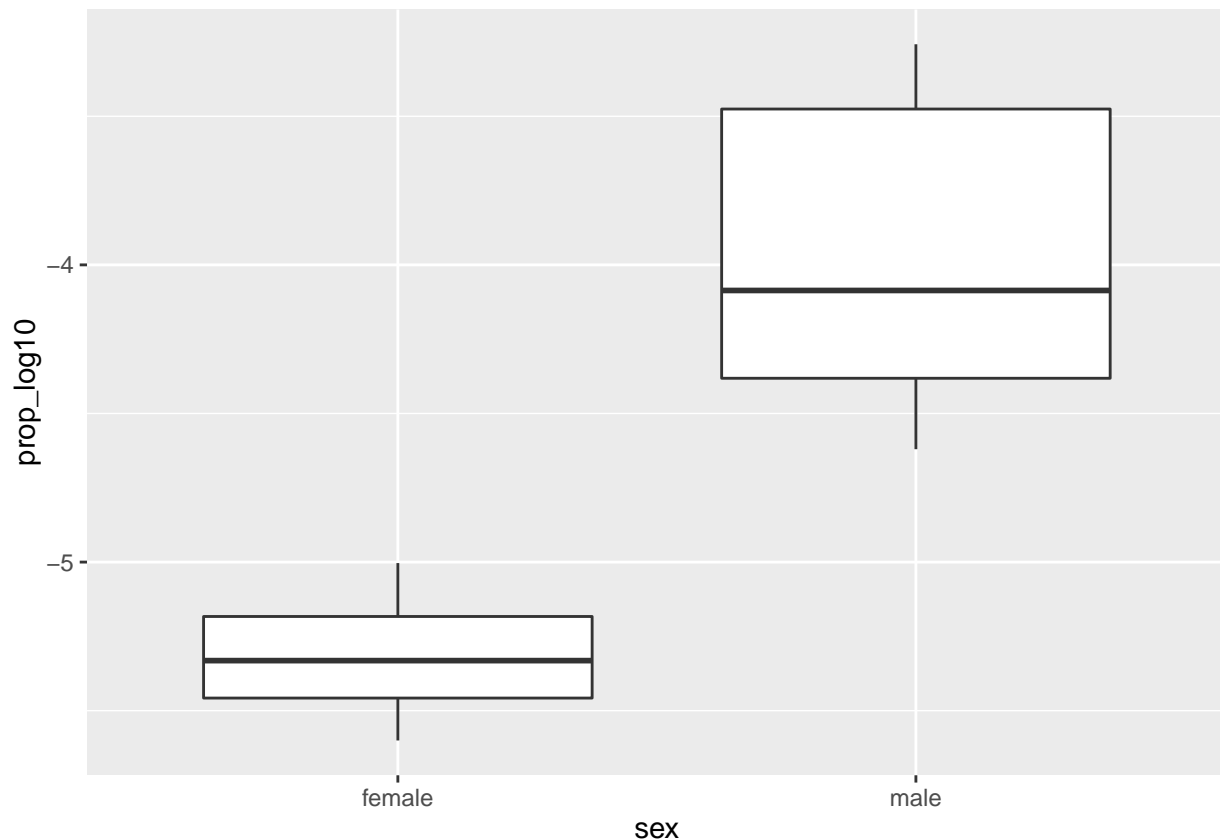
```
year.lm
```

```
##  
## Call:  
## lm(formula = prop_log10 ~ year, data = data_stats)  
##  
## Coefficients:  
## (Intercept)      year  
## -11.193328    0.003544
```

Prevalence by Sex

Below is a plot for the proportion of people with the name “Page” (log base 10 transformed) by sex. It appears there is a clear effect of sex, where it is a less popular name for females than males.

```
sex.plot
```



To test if there is a significant effect of sex a linear model was built. Proportion of the population log base 10 transformed was the dependent variable and sex the independent variable. As show below, sex was significant, with the name being less common in females than males.

```
sex.lm_sum
```

```
##
```

```
## Call:
## lm(formula = prop_log10 ~ sex, data = data_stats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.69780 -0.36989 -0.07417  0.37694  0.66471
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.32512    0.07333  -72.61  <2e-16 ***
## sexM         1.40274    0.08414   16.67  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4083 on 127 degrees of freedom
## Multiple R-squared:  0.6864, Adjusted R-squared:  0.6839
## F-statistic: 278 on 1 and 127 DF, p-value: < 2.2e-16
```

Conclusion

Cedric is more common in males than females.