Your Informative Title Here

Sydney Weisberg and Molly Daniel

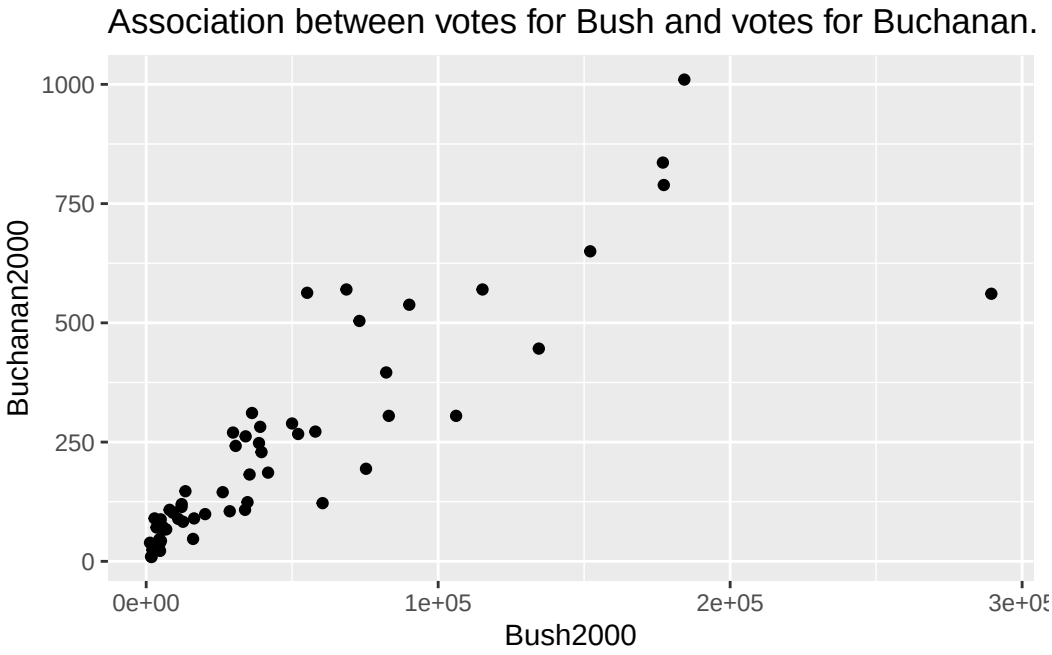
Invalid Date

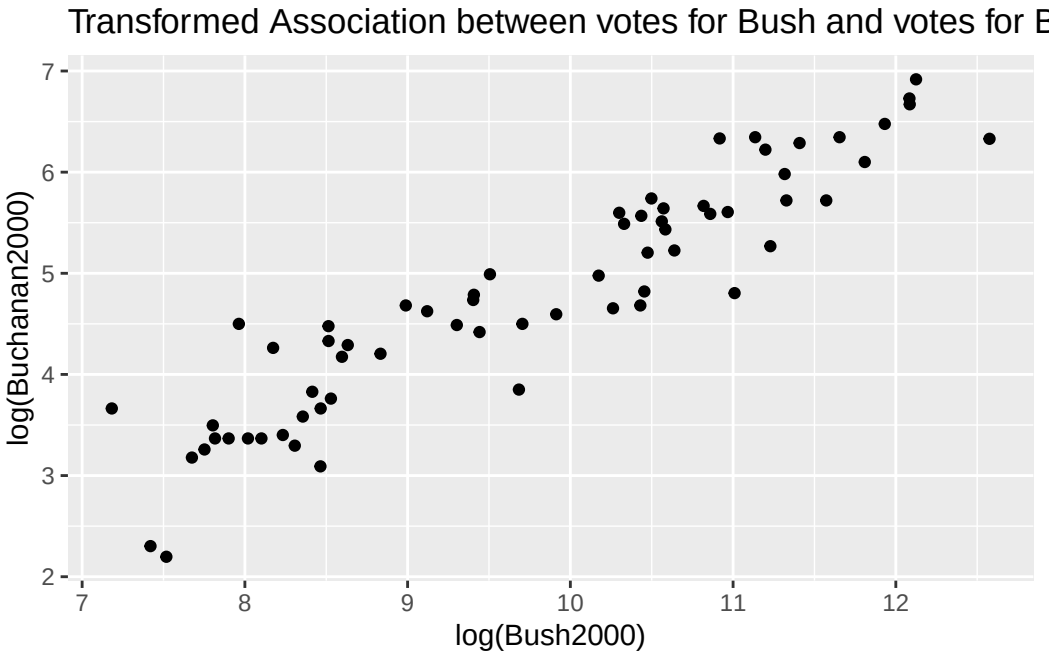
Your written short report should be the first thing your reader encounters in your case study document. While you may choose to label each of the required components/sections of the short report with its own header (see below for an example of creating a header in Quarto), you do not need to do so provided that all of the required information is included. Your report should be self-contained and written in such a way that a quantitatively inclined friend (who has taken or is taking SDS 291) could follow what you did without necessarily knowing anything about the U.S. presidential election of 2000.

As you write your report, you may wish to reference the guide to typesetting regression lines in Quarto using LaTeX (linked at the top of our class Moodle page), the Quarto help page for formatting documents using Markdown (https://quarto.org/docs/authoring/markdown-basics.html), and the Quarto help page for customizing the output from executed code chunks (https://quarto.org/docs/computations/execution-options.html).

An example section heading

When you create plots for your case study report, the `echo: false` *chunk option tells Quarto to include the final output of your R commands (in this case, a plot) in your rendered PDF* without *printing the underlying R commands that generated that plot! The message and warning flags both prevent R from printing any additional text with error messages or warnings to the PDF.*






```
summary(both_transformed)
```

```
Call:
lm(formula = log(Buchanan2000) ~ log(Bush2000), data = election_wo_pb)

Residuals:
     Min       1Q   Median       3Q      Max
-0.95631 -0.21236  0.02503  0.28102  1.02056
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.34149    0.35442  -6.607 9.07e-09 ***
log(Bush2000)  0.73096    0.03597  20.323  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4198 on 64 degrees of freedom
Multiple R-squared:  0.8658,    Adjusted R-squared:  0.8637
F-statistic:   413 on 1 and 64 DF,  p-value: < 2.2e-16
```

Fitted Equation

Let $buchanan_i$ denote the number of votes cast for Buchanan in county i and $bush_i$ denote the number of votes cast for Bush in county i during the U.S. presidential election of November 7, 2000. Our final linear regression model for the mean is $E[log(buchanan_i)|log(bush_i)] = \beta_0 + \beta_1 \log(bush_i)$. We fit our sample data to this model, and found estimates for the coefficients. Our sample intercept, $\widehat{\beta}_0$, is -2.34149 with a standard error of 0.35442. Our sample slope, $\widehat{\beta}_1$, is 0.73096 with a standard error of 0.03597. Both coefficients have a p value less than 0.05, making them statistically significant. So our fitted model is $\log(\widehat{buchanan}_i) = -2.34149 + 0.73096 \log(bush_i)$.

Prediction Interval

```r
library(broom)
predicted_palm_beach = data.frame(Bush2000 = 152846)
both_transformed |> augment(newdata = predicted_palm_beach, interval =
 ↪  "prediction", conf.level = 0.95)
```

```
# A tibble: 1 x 4
  Bush2000 .fitted .lower .upper
     <dbl>   <dbl>  <dbl>  <dbl>
1   152846    6.38   5.52   7.24
```

Point Estimate = e^6.384143 = 592.376848042

Prediction Interval = (e^5.524656, e^7.24363) = (250.8, 1399.164)

This means that we are 95% confident that the number of votes cast for Buchanan in the U.S. presidential election of November 7, 2000 when the number of votes cast for Bush is 152846 (the reported number of votes for Bush in Palm Beach county) is between 250.8 and 1399.164 votes.

When reporting the results of your regression analysis, I would like you to write out a mathematical description of the main/final population model for the mean. Here is an example of what that might look like:

Let $mort_i$ denote the heart disease mortality rate in country i (measured in deaths per 1,000 individuals) and $wine_i$ denote the consumption of wine (measured in liters per person per year). Our final linear regression model is

$$E[mort_i|wine_i] = \beta_0 + \beta_1 (wine_i) .$$

You do not need to write out the fitted regression line, but you should (at a minimum) provide a table summarizing the estimated coefficients of the model and their corresponding standard errors, like the one shown below. A print-out from the `summary()` *command alone is not sufficient.*

	Name for Col. 1	Name for Col. 2	Name for Col. 3	Name for Col. 4
(Intercept)	7.69	0.47	16.24	0e+00
Wine	-0.08	0.02	-4.48	4e-04

R Appendix

Copy and paste all code that you used for your case study into one chunk at the end of your written report. Before submitting your case study, take one final look at the R Appendix and make sure that all code is clearly visible. If you see a line running off the side of the PDF, please split the code over multiple lines using a linebreak. The header at the top of this template should ensure that lines of code are automatically wrapped in your final document.

#| message: FALSE #| warning: FALSE

Loading necessary packages

library(tidyverse) library(Sleuth2) library(broom)
library(kableExtra)

Loading the data for case study one

election <- Sleuth2::ex0825

Creating a second dataset with Palm Beach County excluded

election_wo_pb <- election |> filter(County != “Palm Beach”)

Loading another dataset on wine consumption and heart disease mortality

wine <- Sleuth2::ex0823

Creating a scatterplot for the relationship between mortality and wine consumption

wine |> ggplot(aes(x = Wine, y = Mortality)) + geom_point() + ggtitle("Association between wine consumption and mortality rates.")

Fitting and summarizing the regression line for mean mortality

as a function of wine consumption

wine.lm <- lm(Mortality ~ Wine, data = wine) wine.lm.table <- wine.lm |> tidy() wine.lm.table |> kbl(col.names = c("Name for Col. 1", "Name for Col. 2", "Name for Col. 3", "Name for Col. 4", "Name for Col. 5"), align = "c", booktabs = T, linesep="", digits = c(2, 2, 2, 4)) |> kable_classic(full_width = F, latex_options = c("HOLD_position"))