My first replicable Paper

MyFirstName MyLastName
Evans School of Public Policy and Governance
University of Washington
Seattle, WA 98115, United States
greatguy@uw.edu

January 4, 2018

Abstract

This is an example on how to make a reproducible paper. We are using R from Rstudio, creating an RSweave document. This is a nice start to create a nice paper and get an A+. The next sections will show the steps taken.

1 Introduction

This is my intro to my great paper, I will explain the cool things I can do with my new 'computational thinking' powers combined with some Latex.

This is my nice intro to my great paper, I will explain the cool things I can do with my new 'computational thinking' powers combined with some Latex.

This is my nice intro to my great paper, I will explain the cool things I can do with my new 'computational thinking' powers combined with some Latex.

2 Explaining Labels

Sections may use a label[1]. This label is needed for referencing. For example the next section has label *datas*, so you can reference it by writing: As we see in section 3.

[1] In fact, you can have a label wherever you think a future reference to that content might be needed.

3 Data analysis

Here you can explain how to get the data:

```
> states=read.csv("https://goo.gl/So48s5")
```

3.1 Exploration

Here, I start exploring the data. The first step is to know what variables I have, and in what scale they are:

```
'data.frame':        51 obs. of  8 variables:
 $ state                : Factor w/ 51 levels "Alabama","Alaska",..: 1 2 3 4 5 6 7 8
 $ satMean              : int  991 920 932 1005 897 959 897 892 840 882 ...
 $ satDemand            : num  0.08 0.41 0.26 0.06 0.47 0.29 0.81 0.61 0.71 0.48 ...
 $ k12ExpenditurePupil  : int  3627 8330 4309 3700 4491 5064 7602 5865 9259 5276 ...
 $ incomeHouseholsMedian: num  27.5 48.3 32.1 24.6 41.7 ...
 $ diplomaHsAdults      : num  0.669 0.866 0.787 0.663 0.762 0.844 0.792 0.775 0.731
 $ collegeDegreeAdults  : num  0.157 0.23 0.203 0.133 0.234 0.27 0.272 0.214 0.333 0.
 $ region               : Factor w/ 4 levels "Midwest","N. East",..: 3 4 4 3 4 4 2 3
```

A next step demands:

- Knowing the *central* and *dispersion* values.
- Visualizing the variables of interest.

Except for the column *state* and *region*, we can compute the centrality and spread measures for the other variables in the data. I will do that in Table1 in the next page.

Table 1: Mean and Spread values

Statistic	N	Mean	St. Dev.	Min	Max
satMean	51	944.098	66.935	832	1,093
satDemand	51	0.358	0.262	0.040	0.810
k12ExpenditurePupil	51	5,235.961	1,401.155	2,960	9,259
incomeHouseholsMedian	51	33.957	6.423	23.465	48.618
diplomaHsAdults	51	0.763	0.056	0.643	0.866
collegeDegreeAdults	51	0.200	0.042	0.123	0.333

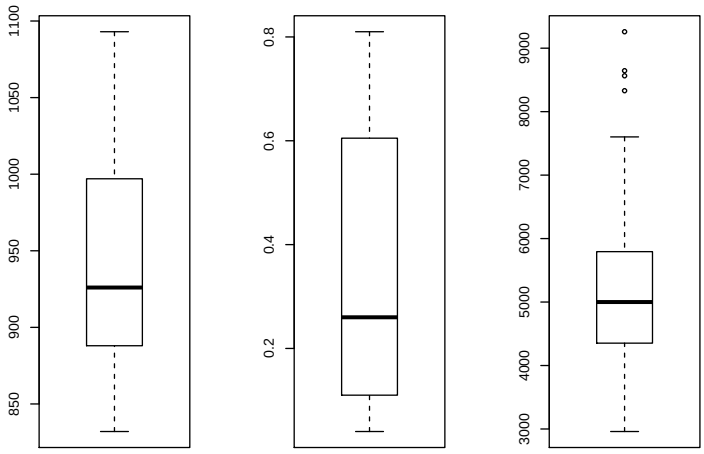

Figure 1: Location of values

As you saw, my Table 1 is nice. As you, saw the mean of the variable *satMean* is 944.1. Now let's use a boxplot to explore location:

As we have a categorical variable, we could create a frequency table:

Table 2: Distribution of Region

Region	Frequency
Midwest	12
N. East	9
South	17
West	13

It does look better now, right? let's work on testing some hypothesis.

3.2 Modeling

Here, I propose that the amount of money spent for child per state in the US has an effect on the mean average pupils in a state get in SAT:

```
> reg1=lm(satMean~k12ExpenditurePupil, data = states)
> summary(reg1)

Call:
lm(formula = satMean ~ k12ExpenditurePupil, data = states)

Residuals:
     Min       1Q   Median       3Q      Max
-131.811  -38.085    5.607   37.852  136.495

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.061e+03  3.270e+01   32.44  < 2e-16 ***
k12ExpenditurePupil -2.228e-02  6.037e-03   -3.69 0.000563 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59.81 on 49 degrees of freedom
Multiple R-squared:  0.2174,        Adjusted R-squared:  0.2015
F-statistic: 13.61 on 1 and 49 DF,  p-value: 0.0005631
```

Here, I modify the previous model; while I insist that the amount of money spent for child per state in the US has an effect on the mean average pupils in a state get in SAT; I will control the effect the demand per state (as demand were equal accross states). Then,

```
> reg2=lm(satMean~k12ExpenditurePupil+satDemand, data = states)
> summary(reg2)

Call:
lm(formula = satMean ~ k12ExpenditurePupil + satDemand, data = states)

Residuals:
    Min      1Q  Median      3Q     Max
-62.921 -24.318   1.741  15.502  75.623

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          9.898e+02  1.840e+01  53.806  < 2e-16 ***
k12ExpenditurePupil  8.604e-03  4.204e-03   2.046   0.0462 *
satDemand           -2.538e+02  2.249e+01 -11.283 4.21e-15 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31.62 on 48 degrees of freedom
Multiple R-squared:  0.7857,        Adjusted R-squared:  0.7768
F-statistic: 88.01 on 2 and 48 DF,  p-value: < 2.2e-16
```

Do you like the way my results are shown?...I will improve them later.

Try 'uncommenting' the re encoding package *inputenc* on this document header.