Example for knitting pdf with models

Introduction

This is a causal analysis on fourth graders in Massachusetts public schools in the spring of 1998. We investigating the effect of student per teacher ratio on the test results of the fourth graders. ECT.

HERE COMES THE MOTIVATION WHY THIS IS A MEANINGFUL PROJECT AND WHAT IS THE MAIN GOAL!

Data

The Massachusetts data are ... Further information is available (here)[https://www.rdocumentation.org/packages/AER/versions/1.2-9/topics/MASchools].

ECT.

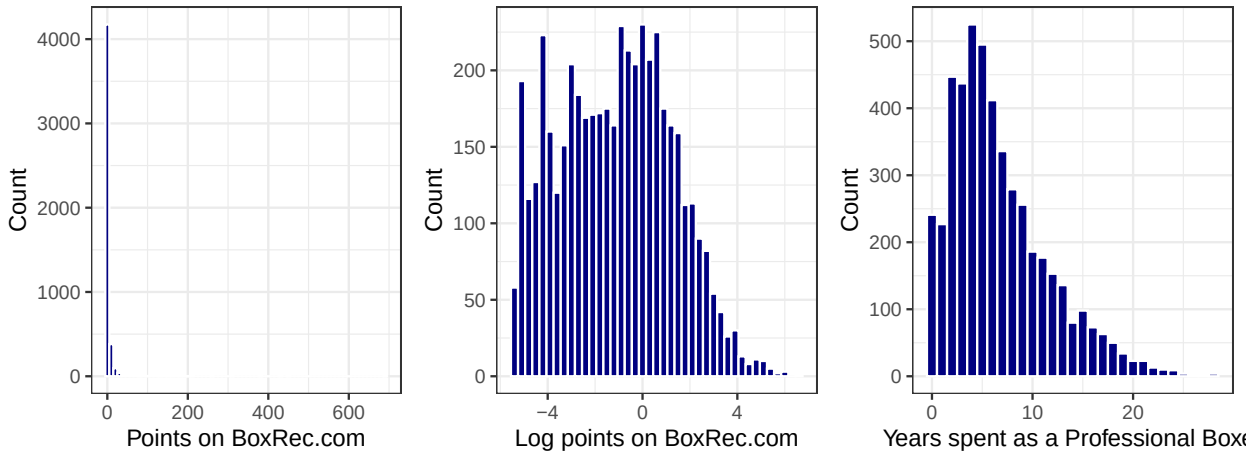
Table 1: Descriptive statistics

	Mean	Median	SD	Min	Max	P05	P95
Rank on BoxRec.com	4153.28	3691.00	2888.10	1.00	9997.00	274.70	9276.60
Points on BoxRec.com	4.11	0.36	20.69	0.00	691.00	0.01	14.97
Age	29.27	29.00	5.64	15.00	52.00	21.00	39.00
Number of professional bouts	18.64	15.00	15.07	1.00	248.00	3.00	47.00
Years spent as a professional boxer	6.70	6.00	4.87	0.00	28.00	0.00	16.00
Age of becoming a professional boxer	22.56	22.00	3.98	12.00	44.00	17.00	30.00
Height to average height per division ratio (%)	100.00	100.00	2.79	86.00	114.00	95.00	104.00

The number of observations is 4795 for all of our key variables.

DESCRIPTION OF THE SUMMARY STATS: WHAT CAN WE LEARN FROM THEM?

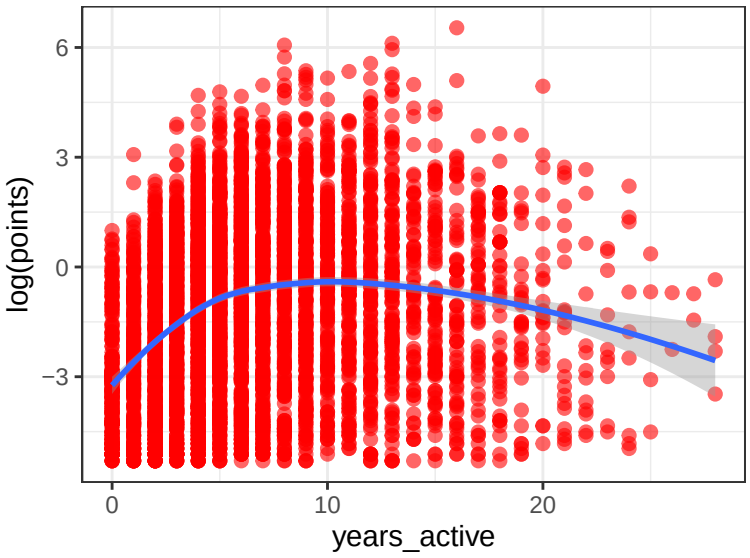
As the focus is the price difference, the next Figure shows the histogram for this variable.



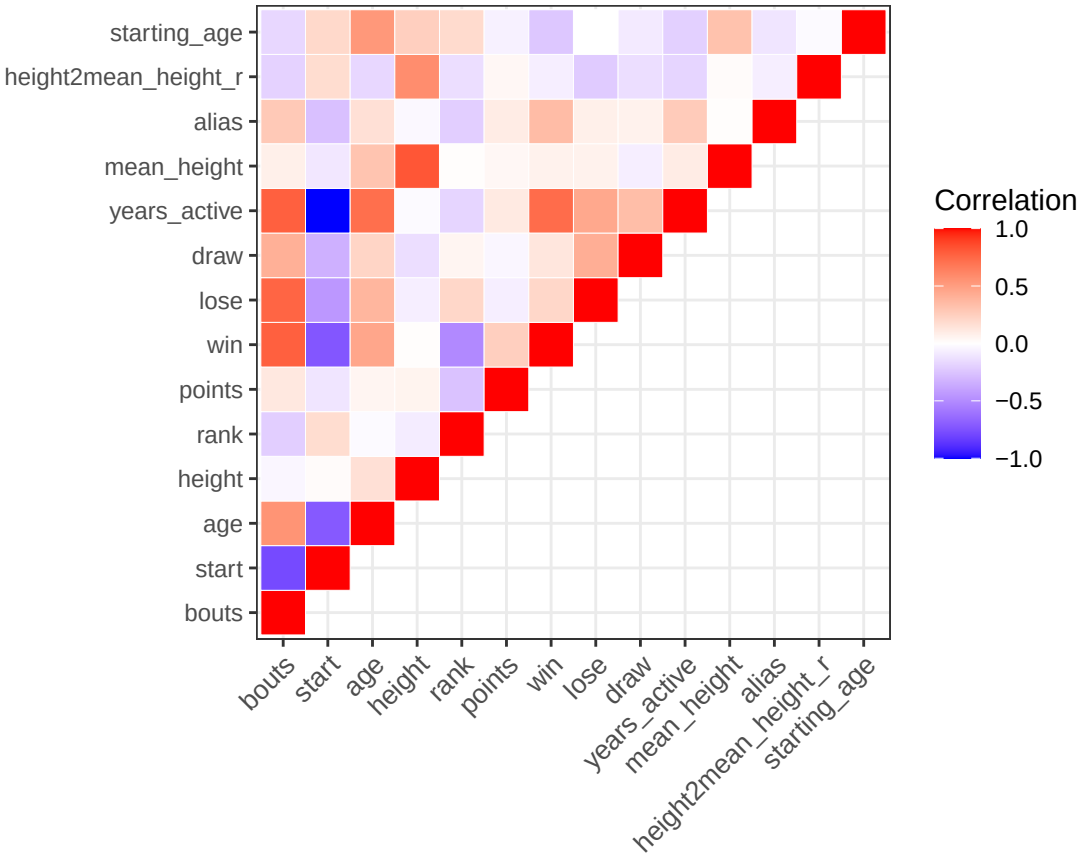
DESCRIPTION OF THE FIGURE. WHAT DOES IT TELS US?

(May change the order of descriptive stats and graph.)

The key pattern of association is:



Heatmap



How will you include this in your model?

Short description on the other variables: 2-10 sentence depends on the amount of variables you have. You should reference your decisions on the graphs/analysis which are located in the appendix.

Model

```
## The variable 'lspline(years_active, 5)2' has been removed because of collinearity (see $collin.var).
```

My preferred model is:

$\log(\text{points}) = -14.32\ 0.23\ (bouts < 25)\ 0.14\ (bouts \geq 25) + \delta Z$

where Z are standing for the controls, which includes controlling for height to mean height per division, number of losses, current age, starting age, whether the boxer has an alias on BoxRec.com and active years. From this model we can infer:

- In case of log-level regression, the intercept is practically meaningless (when the boxer had 0 professional bouts, he would be expected to have a point of -14.32)
- when the number of bouts is one unit larger, but below the value of 25, we see boxers to have on average 23 % more points
- when the number of bouts is one unit larger, with the value above or equal to 25, boxers to have on average 14 % more points.

Based on the heteroskedastic robust standard errors, these results are statistically different from zero. To show that, I have run a two-sided hypothesis test:

$$H_0 := \beta_1 = 0$$

$$H_A := \beta_1 \neq 0$$

I have the t-statistic as 31.4 and the p-value as 0 for when the number of bouts is less than 25, and 14.22 and the p-value as 0 for when the number of bouts is more than 25, which confirms my conclusion.

We compare multiple models to learn about the stability of the parameters. Bla-bla:

[H]

Table 2: Models to uncover relation between test score and student to teacher ratio

	(1)	(2)	(3)	(4)	(5)
Intercept	-1.798***	-3.266***	-3.325***	-14.35***	-14.32***
	(0.0727)	(0.0629)	(0.0639)	(1.146)	(1.080)
bouts	0.0347***				
	(0.0038)				
bouts (<25)		0.1523***	0.1431***	0.1811***	0.2280***
		(0.0047)	(0.0048)	(0.0084)	(0.0073)
bouts (>=25)		-0.0542***	-0.0546***	0.0812***	0.1387***
		(0.0053)	(0.0054)	(0.0079)	(0.0098)
alias_dummy			0.4417***	0.1040	0.0613
			(0.0662)	(0.0566)	(0.0530)
height to division mean height ratio				0.1110***	0.0917***
				(0.0116)	(0.0100)
years active (<5)					0.2061***
					(0.0338)
Losses	No	No	No	Yes	Yes
Age	No	No	No	No	No
Observations	4,795	4,795	4,795	4,795	4,795
R2	0.04785	0.20522	0.21272	0.42041	0.49011

Robustness check / 'Heterogeneity analysis'

Task: calculate and report t-tests for each countries.

Conclusion

HERE COMES WHAT WE HAVE LEARNED AND WHAT WOULD STRENGHTEN AND WEAKEN OUR ANALYSIS.

Appendix

Here comes all the results which are referenced and not essential for understanding the MAIN results.