Getting to know my audience

An example of literate programming

October 7, 2016

Introduction

This report analyzes survey data collected at the Emphasizing Reproducible Research by Teaching Data Analysis with R Presentation at the CELT 2016 Conference. After setting up proper authorization tokens, we connect directly to Google Sheets to download the data.

Looking at the raw data, we see that the names of the data frame are set to the question titles, which makes analysis very difficult to deal with. We rename these variable names something easier to manage.

Description of the data set

College

AGR	BSS	BUS	CME	ECC	HFA	NS
2	5	2	1	1	2	4

Figure 1 shows that the majority of the audience (n=5 (29.4)%) is from the college of BSS.

Usage of R

	No	Yes	Sum
No	7	2	9
Yes	4	4	8
Sum	11	6	17

There are 10 (58.8%) people in the audience who use R for their teaching or research. Of the 6 who use it for research, 2 (33.3%) don't use it for teaching.

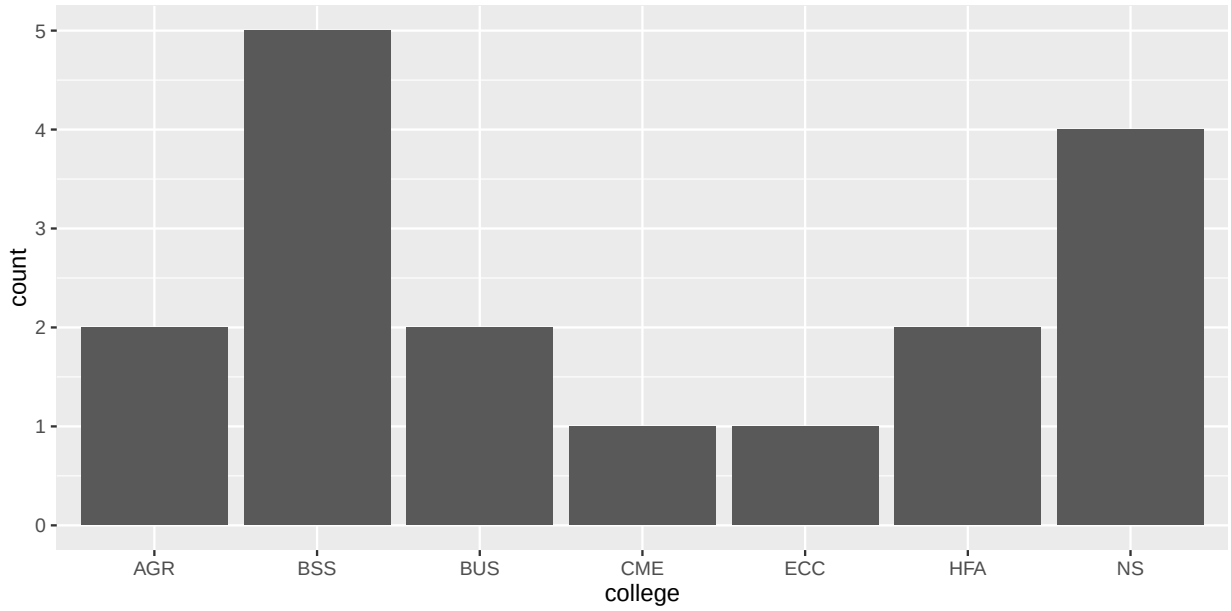
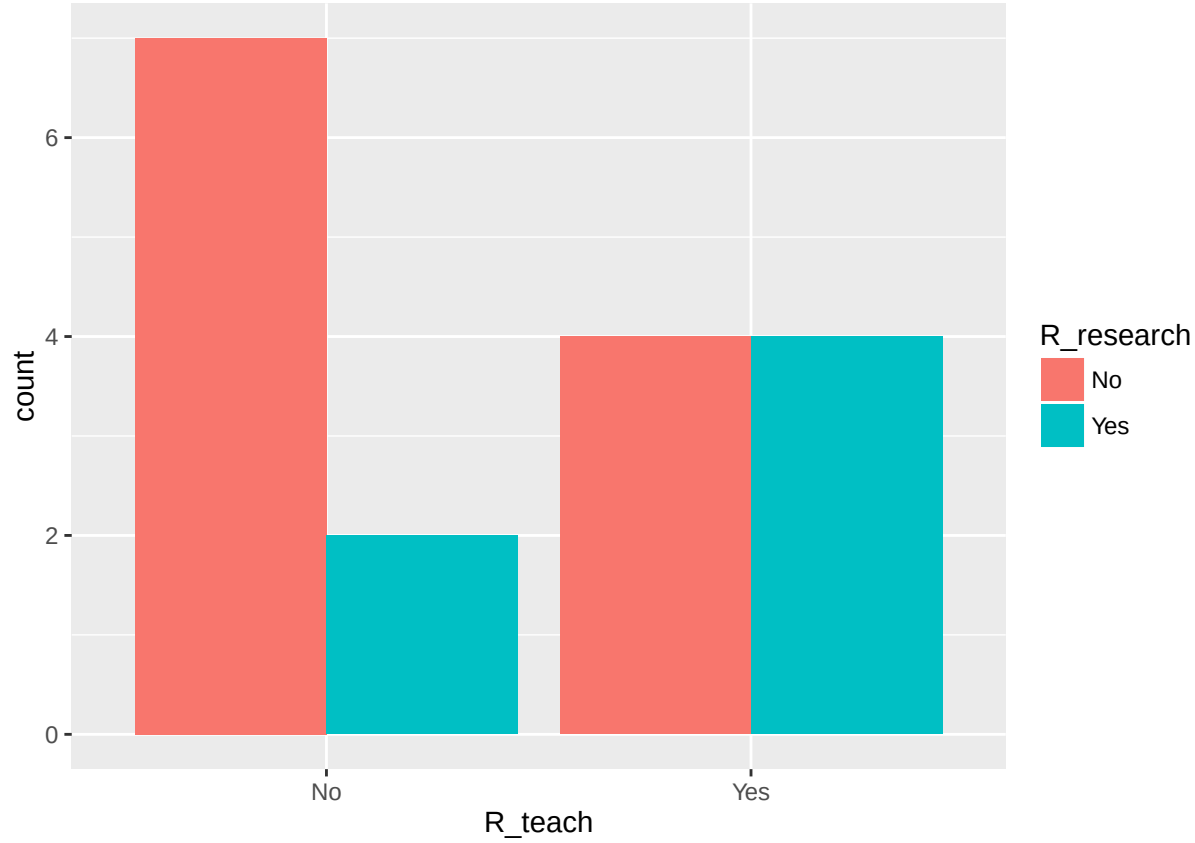
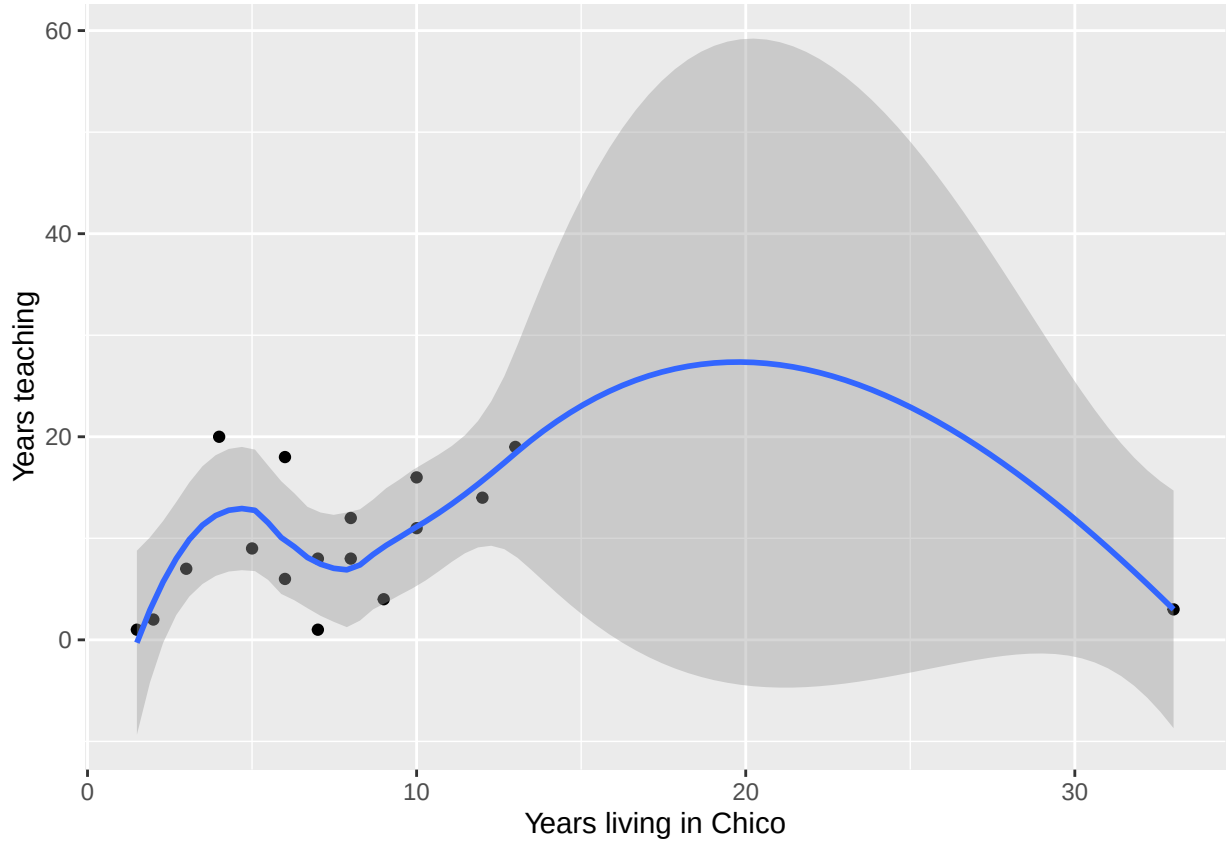


Figure 1: College Representation

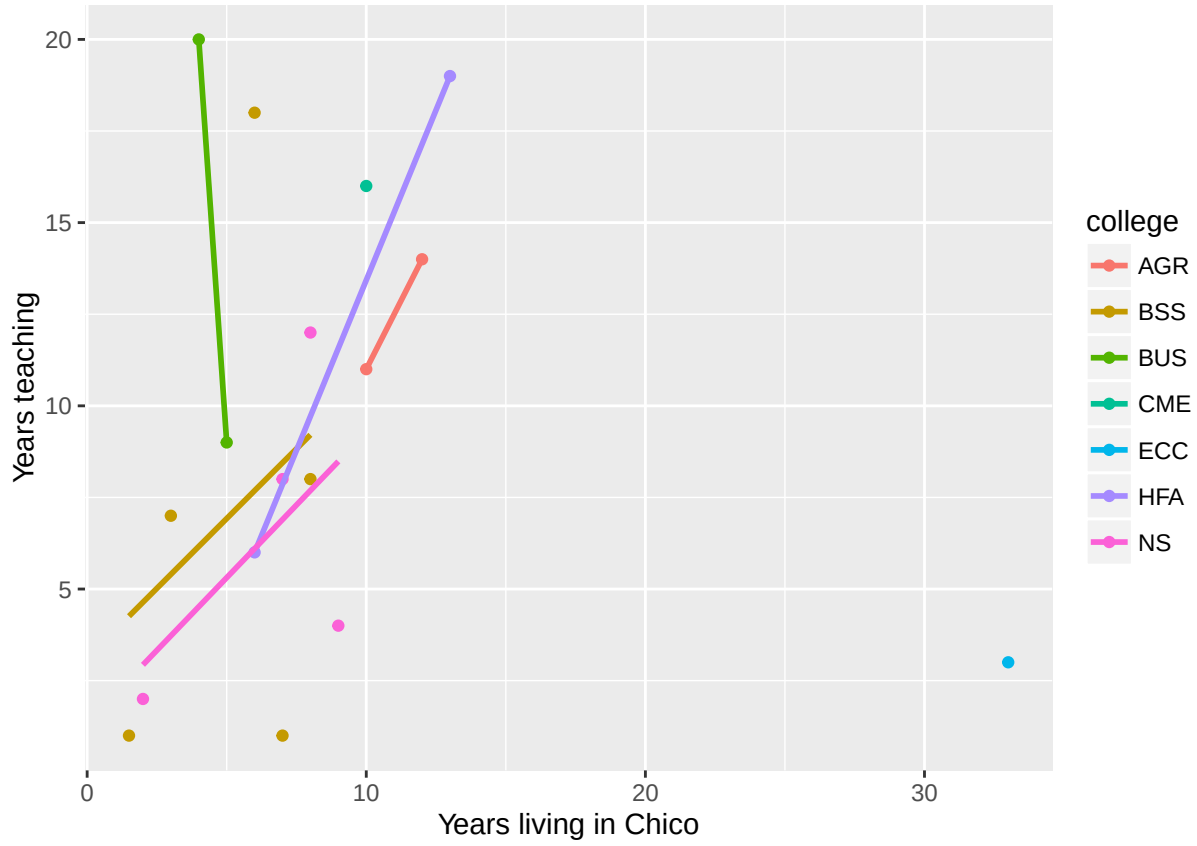


Years teaching

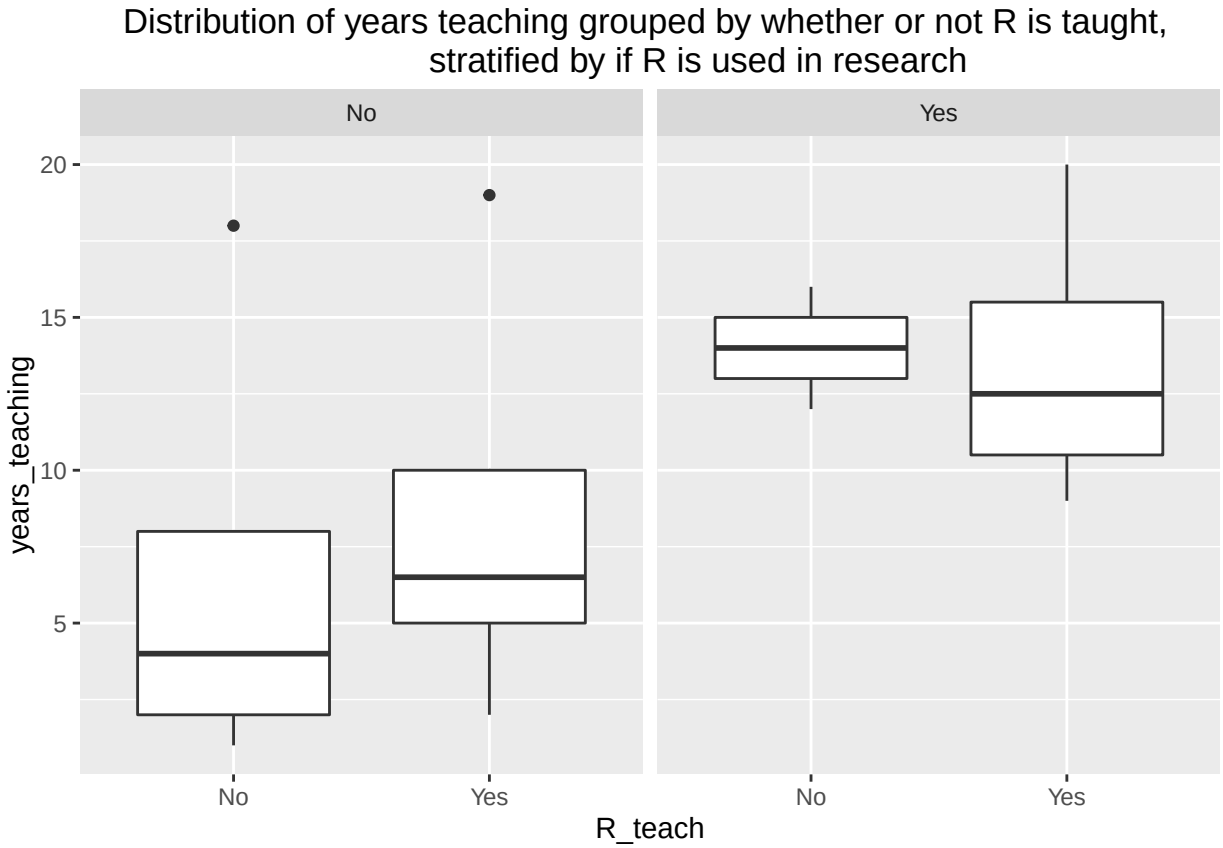
The correlation between years teaching overall and years living in Chico is -0.02. The plot below demonstrates that this relationship is *positive*.



Does this relationship differ by college?



Now the big question: Is the use of R in research or teaching a relative new thing?



An ANOVA was used to test if there is a significant difference in the average years teaching for those who teach, or do research in R.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
R_teach	1	40.99	40.99	1.25	0.28
R_research	1	138.89	138.89	4.23	0.06
Residuals	14	460.00	32.86	NA	NA

The results show that there is a non-significant difference in the average age of those who teach R and those who do not, and a non-significant difference between those who do research with R and those who do not.