

- 1 Background
- 2 Study Population
- 3 Methods
- 4 Descriptive Statistics
- 5 Notes
- 6 Questions

Metadata Exploratory Report

Code ▼

Project: Differential Gene Expression and Methylation in Vaping LatinX Youth

Analyst: ~~Cheyret Wood~~ ; Trent Hawkins

Investigator(s): Sunita Sharma, MD

Report generated: February 28, 2022

Center for Innovative Design & Analysis

colorado school of public health

1 Background

The data presented in this report are part of a study aimed to assess differential gene expression and methylation in vaping versus non-vaping LatinX youths in Pueblo and Denver, CO. Pulmonary function data were also obtained in order to better understand the impacts of vape use on pulmonary function. To assess differential gene expression and methylation, naso-epithelial swabs were obtained from each participating subject. Pulmonary function is assessed using PFTs (Pulmonary Function Tests) and Impulse Oscillometry (IOS).

2 Study Population

This data set consists of samples taken from 51 people ages 12-17 from the Pueblo, Denver, and Aurora, CO areas. Subjects were asked to identify as 'LatinX' or 'Non-latinX'.

3 Methods

[Include the version of R you are using.](#)

3.1 Clinical Data Pocessing

This is a great summary of what you've done, it includes the important bits and describes how you used the other variables.

- Subjects are dichotomized to those that used a vaping device in the last 6 months and those who have not based on the variables 'ever_vape', 'vape_days', and 'last_vape'. *Six Month Vape Status* could not be confirmed for one individual. Any analyses including *Six Month Vape Status* will have a sample size of n = 50 assuming no other missing values. Vaping status is self-reported. Previous analyses showed n = 12 participants had vaped in the last 6 months. This analysis will use n = 13 participants who had vaped in the last 6 months. One participant (SID = 111) reported that they had used a vaping device 5 out of the last 30 days, but did not respond to 'last_vape'. They were ~~falsely~~ previously labeled as "NA" in previous analyses.
- Subjects' geographic location (City) is grouped by their reporting recruiting center.

So we will also want a sentence about our lung function measures. It might just be that normality was visually inspected using histograms, but we might not have a normality assumption for this measures to pass.

3.2 Gene-Count Processing

Key Pts: versions of annotation (ensembl in this case), version of packages (DESeq2 and RUVSeq)

- Library Size:
 - Example: Gene counts were quantified using Ensembl annotation for GrCh 38 ver. 86 (pulled 08/23/2017). Genes were filtered to only include those with an average of 10 reads per sample.
- Normalization:
- Technical effects:

NORMALIZATION BIT: To correct for batch and other technical effects Removal of Unwanted Variance with empirically derived control gene (RUVg) was used. Empirical genes were identified by first making the data Gaussian by performing a variance stabilizing transformation (VST), and then subsetting the quartile of least variable genes across all samples. This was 3,992 genes or 25% of 15,970. 3 nuisance factors were chosen to be removed using this method (k=3).
- Normalization:

TRANSFORMATION: Gene expression values were VST using the R package DESeq2 ver 1.28.1. (VST is equivalent to Rlog when size estimates vary little between subjects as was in this case).

4 Descriptive Statistics

Just relabel as results

Aim to always start with a sentence. Here just state the number of subjects you have after cleaning the data and if any outliers were detected (which no).

Table 1: Clinical Data

Code

	Did Not Vape in Last 6 Months (N=37)	Vaped in Last 6 Months (N=13)	Total (N=50)	p value
Gender	I would prefer to include sex over gender, we are concerned more about accounting for biological differences, not necessarily gender identity differences			0.475 ¹
Female	20 (54.1%)	5 (38.5%)	25 (50.0%)	
Male	16 (43.2%)	8 (61.5%)	24 (48.0%)	
Non-Binary	1 (2.7%)	0 (0.0%)	1 (2.0%)	
Age	Include units so Age (yrs)			0.663 ²
Mean (SD)	14.6 (1.4)	14.8 (1.4)	14.6 (1.4)	
Range	12.0 - 17.0	13.0 - 17.0	12.0 - 17.0	
Grade				0.457 ³
7th	3 (8.1%)	0 (0.0%)	3 (6.0%)	
8th	8 (21.6%)	4 (30.8%)	12 (24.0%)	
Freshman	14 (37.8%)	3 (23.1%)	17 (34.0%)	
Sophomore	4 (10.8%)	2 (15.4%)	6 (12.0%)	
Junior	4 (10.8%)	2 (15.4%)	6 (12.0%)	
Senior	4 (10.8%)	2 (15.4%)	6 (12.0%)	

Use 9th, 10th, 11th and 12th
Freshman and sophomore ect are also
used in college so while it's obvious better
to be clear

	Did Not Vape in Last 6 Months (N=37)	Vaped in Last 6 Months (N=13)	Total (N=50)	p value
City <small>Maybe since you grouped based on reporting recruitment center, put recruiting center over city here</small>				< 0.001 ¹
Aurora	16 (43.2%)	0 (0.0%)	16 (32.0%)	
CommCity/Denver	11 (29.7%)	1 (7.7%)	12 (24.0%)	
Pueblo	10 (27.0%)	12 (92.3%)	22 (44.0%)	
Ethnicity				0.135 ¹
LatinX	23 (62.2%)	11 (84.6%)	34 (68.0%)	
Non-LatinX	14 (37.8%)	2 (15.4%)	16 (32.0%)	
FEV1/FVC <small>Include FEV1 as its own measure as well</small>				0.494 ²
N-Miss	10	12	22	
Mean (SD)	0.8 (0.1)	0.7 (NA)	0.8 (0.1)	
Range	0.5 - 1.0	0.7 - 0.7	0.5 - 1.0	
R5				0.007 ²
N-Miss	1	0	1	
Mean (SD)	4.0 (0.9)	5.0 (1.3)	4.3 (1.1)	
Range	2.0 - 6.1	3.7 - 7.6	2.0 - 7.6	
X20				0.007 ²
N-Miss	4	2	6	
Mean (SD)	0.1 (0.6)	0.7 (0.9)	0.2 (0.7)	
Range	-1.1 - 2.4	-1.0 - 2.3	-1.1 - 2.4	

1. Pearson's Chi-squared test
2. Linear Model ANOVA
3. Trend test for ordinal variables

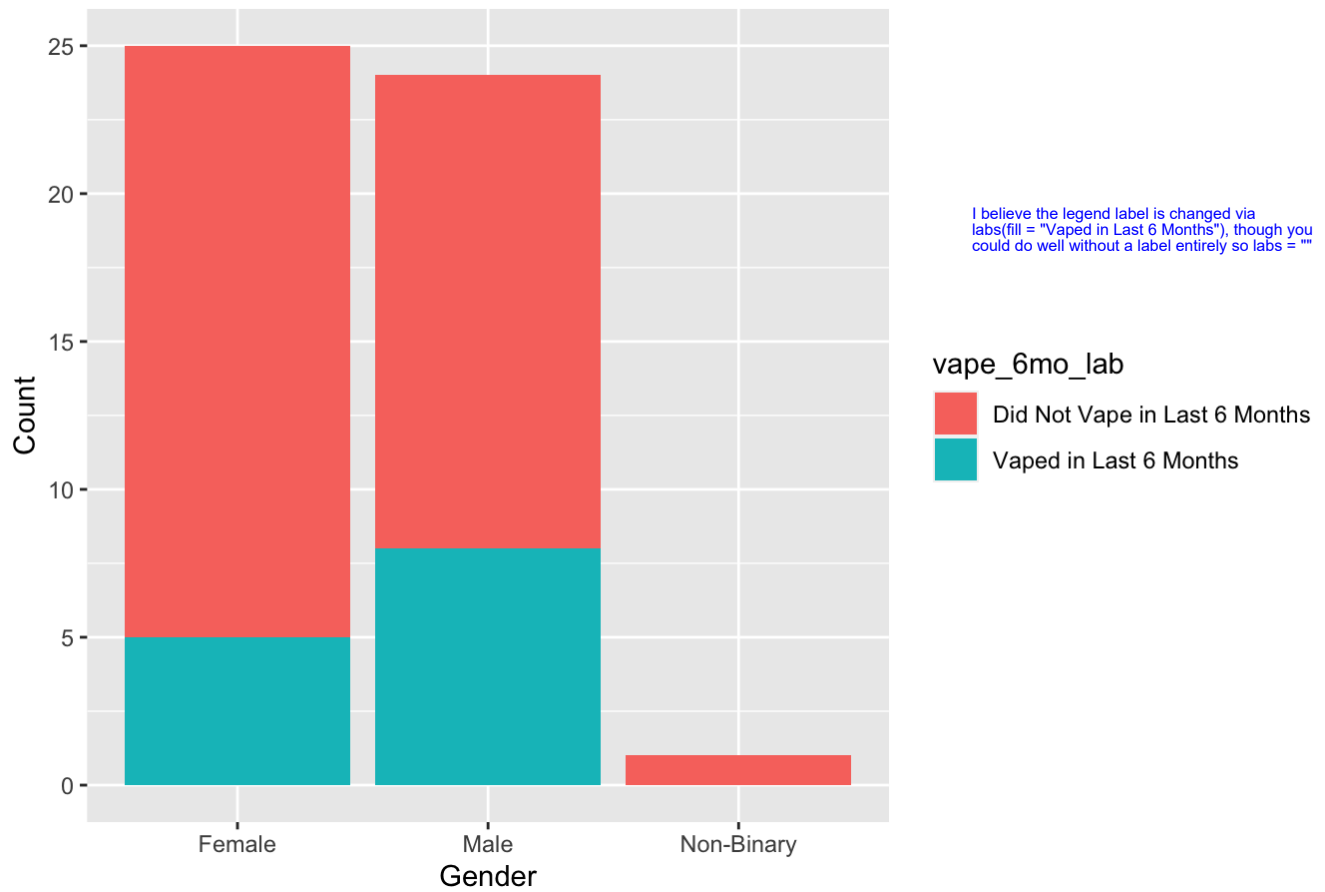
4.1 Figures

So figures would just be built into the results, so this doesn't really need a header. And again just add a sentence here, you could talk about which of the values above were significantly different between our groups and if you can what that means.

Code

After looking at the figures you have, i think grouping some of these together into grids could help especially if you want to make a statement about them together. I would probably group demographic info and lung function info together

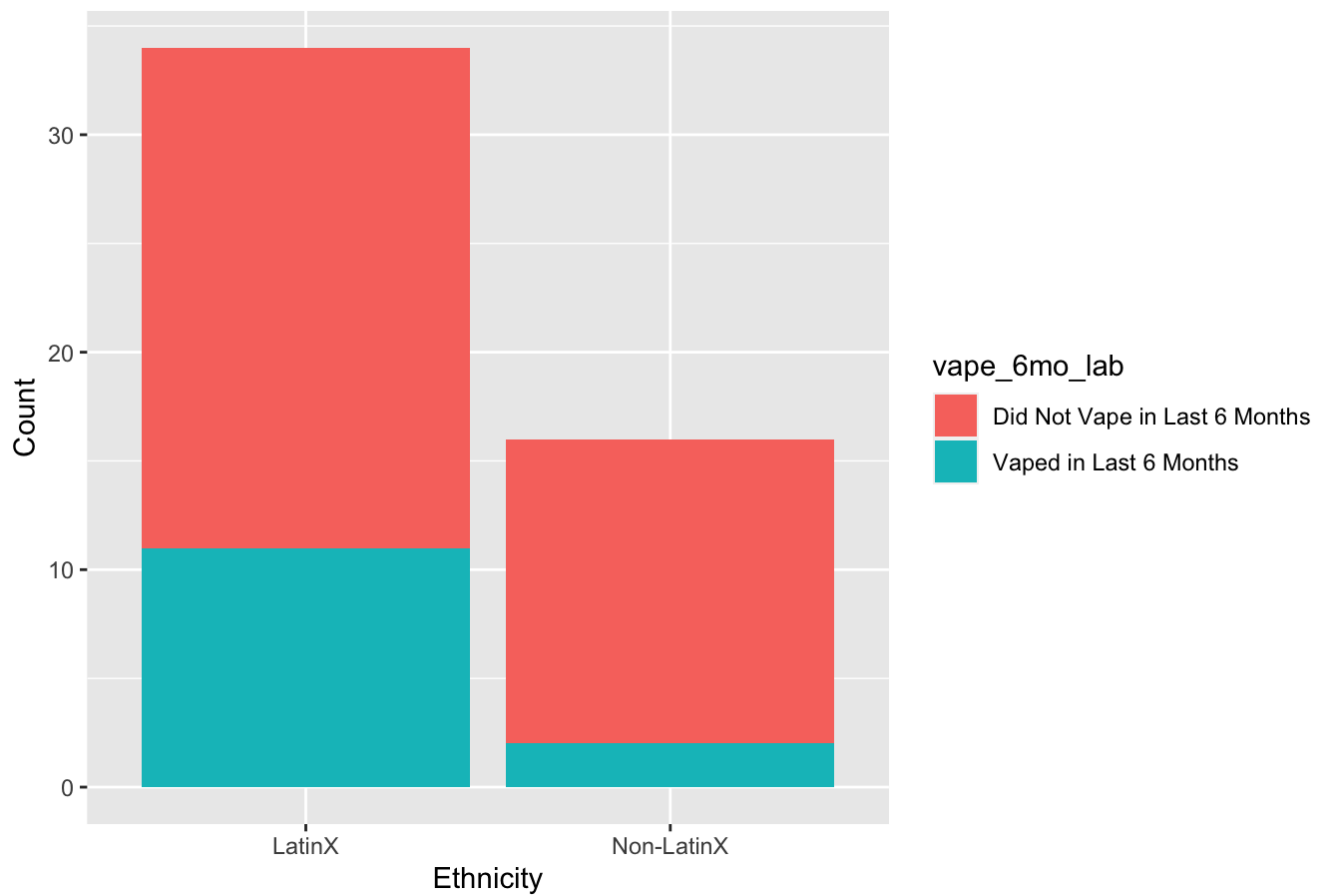
Vaped in the last 6 months by Gender (n = 50)



Code

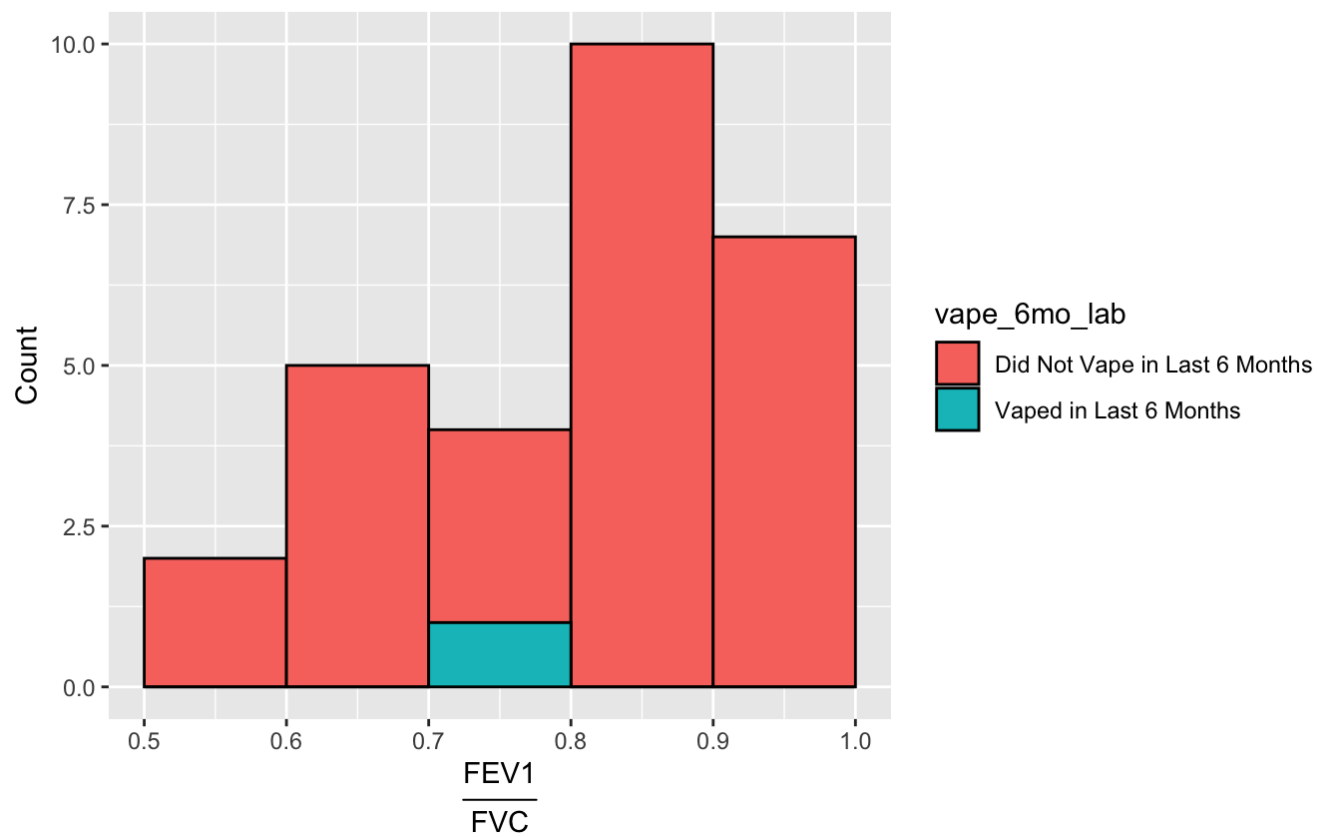
Try to have a sentence for each figure. It helps justify why you have it. If you can't find a sentence to justify maybe it needs some additional information added to it

Vaped in the last 6 months by Latino (n = 50)



Code

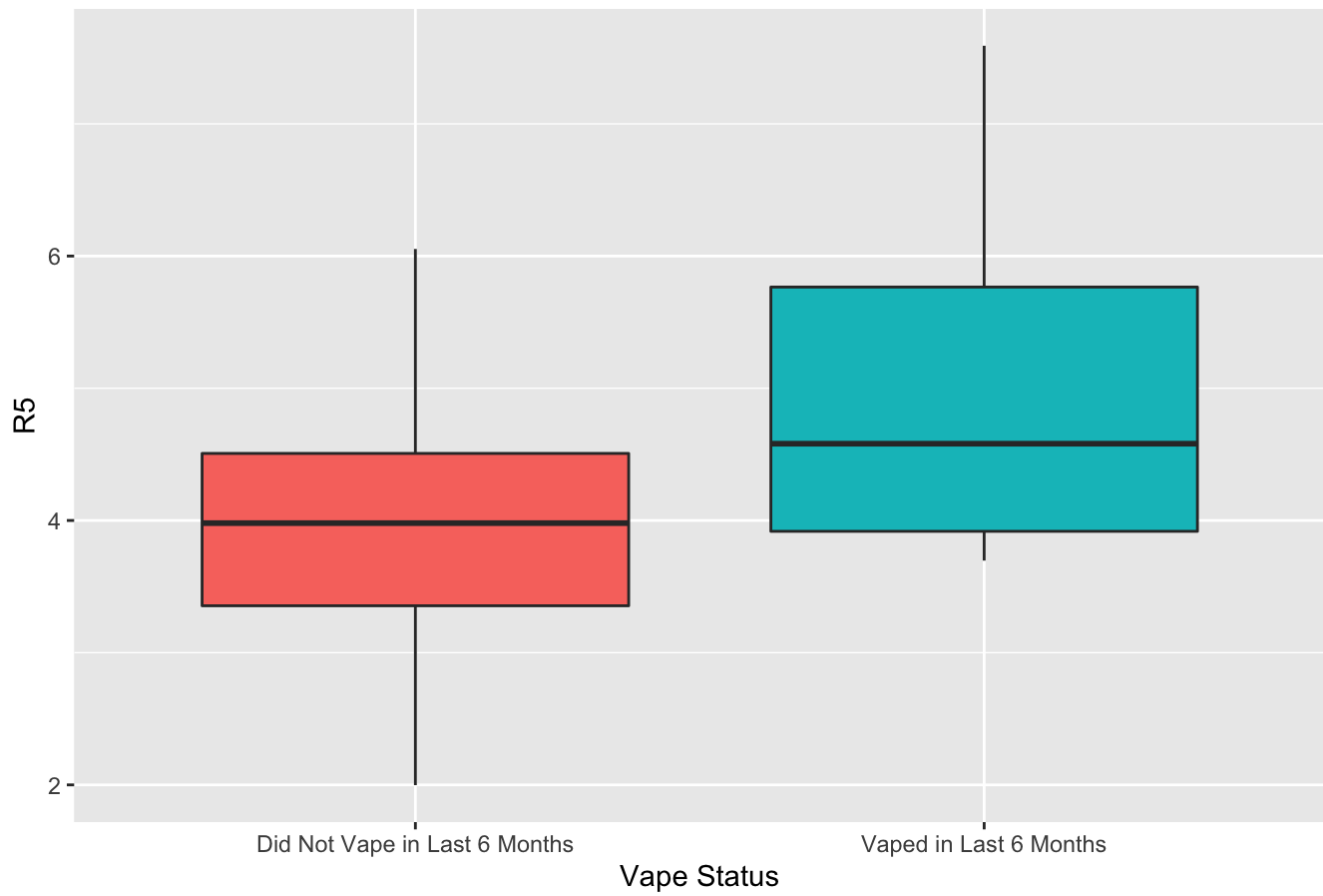
$\frac{FEV1}{FVC}$ by Vape Status (n = 28)



Code

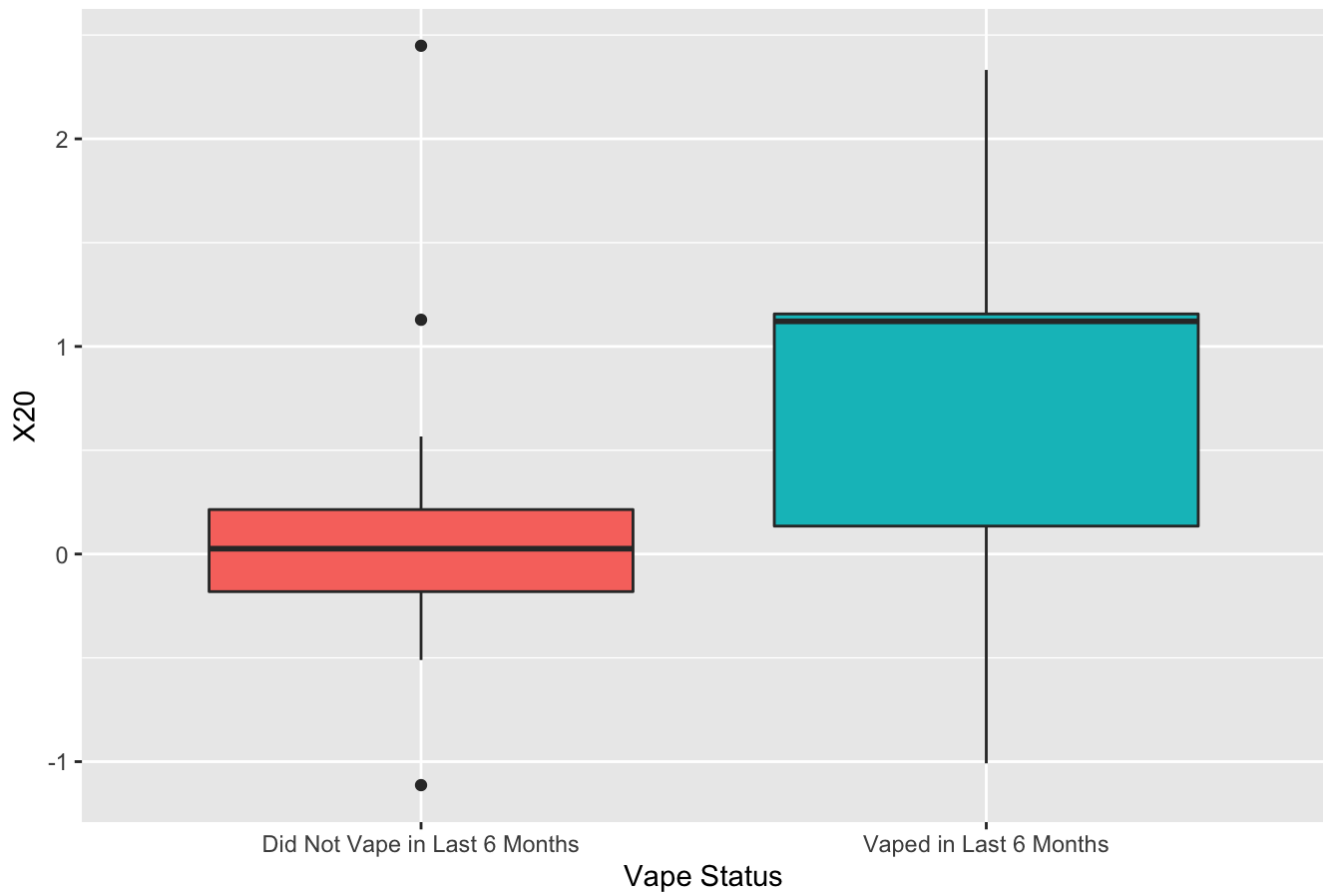
R5 by Vape Status (n = 49)

You could also do color coded histograms with an alpha = 0.8 or something, but a boxplot also works



Code

X20 by Vape Status (n = 44)



5 Notes

6 Questions

Center for Innovative Design & Analysis

colorado school of public health