Lab 1

Matt Massey

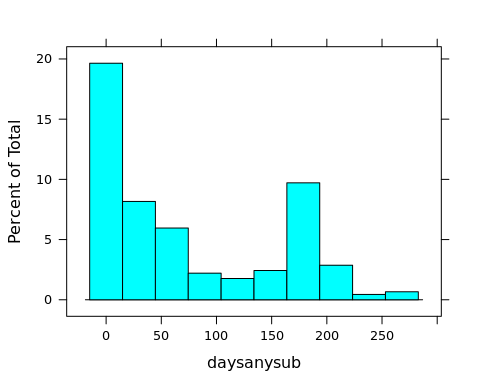
5/20/2021

Question 5. The HELPrct dataset is based on the HELP study, which was a clinical trial of recruited adult inpatients admitted for detoxification and who had no primary care physician. These patients were randomly selected to receive a multidisciplinary assessment and a brief motivational intervention or usual care. The goal of this study was to link treatment to primary medical care.

Question 6. Three variables were chosen for this lab: substance, which is a categorical variable of primay abused substance by the patient; avg\_drinks, which is a quantitative variable of the average number of alcoholic drinks per day by the patient; daysanysub, which is a quantitative variable of the number of days to the first post-detox. substance use.

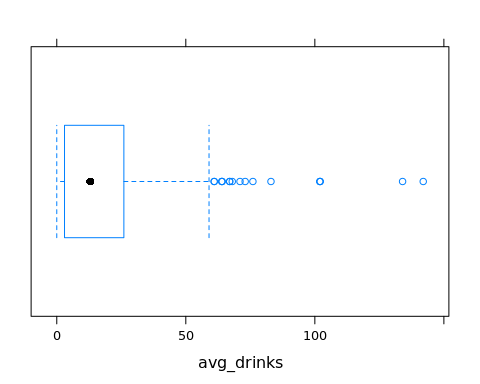
Question 7.

histogram(~daysanysub, data=HELPrct, type="percent")



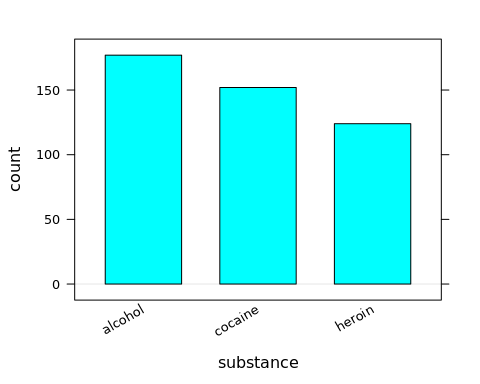
Question 7b. The histogram of daysanysub is bimodal with a pimary mode at 0 days and a subordinate mode at ~180 days. Overall the data is skewed to the right (positive skew) and shows about ~280 days spread.

bwplot(~avg\_drinks, data=HELPrct)



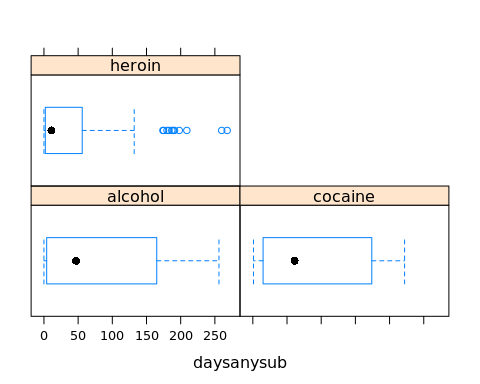
Question 7b. The modified boxplot of avg\_drinks shows a spread from 0 drinks to almost 150 drinks, a relatively narrow IQR ~25, a median of ~10, and a number of outliers beyond ~60 drinks. Overall the data is skewed to the right (positive skew).

bargraph(~substance, data=HELPrct)



Question 8.

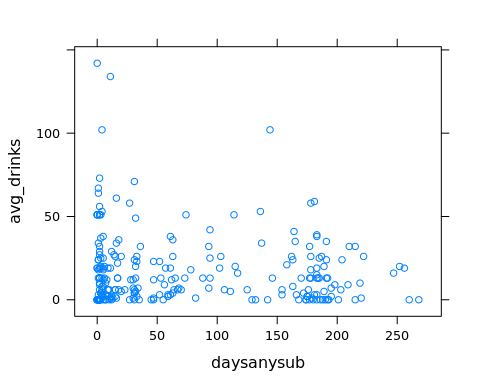
bwplot(~daysanysub | substance, data=HELPrct)



The three comparative modified boxplots above show daysanysub by substance. The median and IQR for both alcohol and cocaine are very similar at ~50 days and ~150 days, respectively, while the total spread is slightly lower for cocaine (~225 days) vs. alcohol (~255 days). Neither alocohol nor cocaine show any outliers. In contrast, heroin shows a much lower median (~5 days), IQR (~50 days), and total spread (~140 days); heroin also shows a number of outliers beyond 150 days. Alcohol, cocain, and heroin are all skewed to the right to varying degrees, with cocaine the least (close to symmetrical), followed by alcohol, and then heroin.

Question 9.

xyplot(avg\_drinks~daysanysub, data=HELPrct)



There doesn’t appear to be any obvious relationship between avg\_drinks and daysanysub. People with avg\_drinks above ~60 seem have mostly lower daysanysub, however, this is such a small number of samples I don’t think this is conclusive. There are also two weak clusters of data points at/near 0 daysanysub and 180 daysanysub, which is also shown in the histogram above.

Question 10. Several questions I am interested in from the data above:

What is the signficance of the bimodal distribution of daysanysub at 0 and 180? Is this real or from the sampling method?

What is different about the heroin outliers (boxplot)? Are these truly outliers? If so, what is different about these patients that allowed them to not use any substances for so long compared to the other heroin patients?