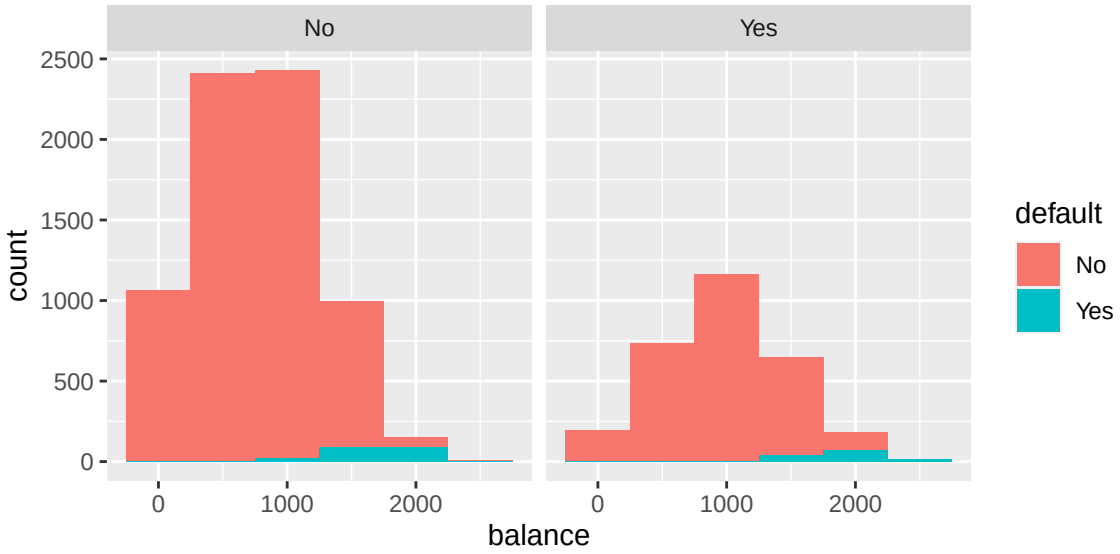
C1b___Homework_EDA

Credit Card Data Analysis

Using the Credit Card data we've been working with, use the following code to visualize the distribution of defaults across balances and student dimensions:

```
dfDefault = Default

p1 = ggplot(data = dfDefault, aes(balance, fill = default)) +
  geom_histogram(binwidth = 500) +
  facet_wrap(~ student)
p1
```

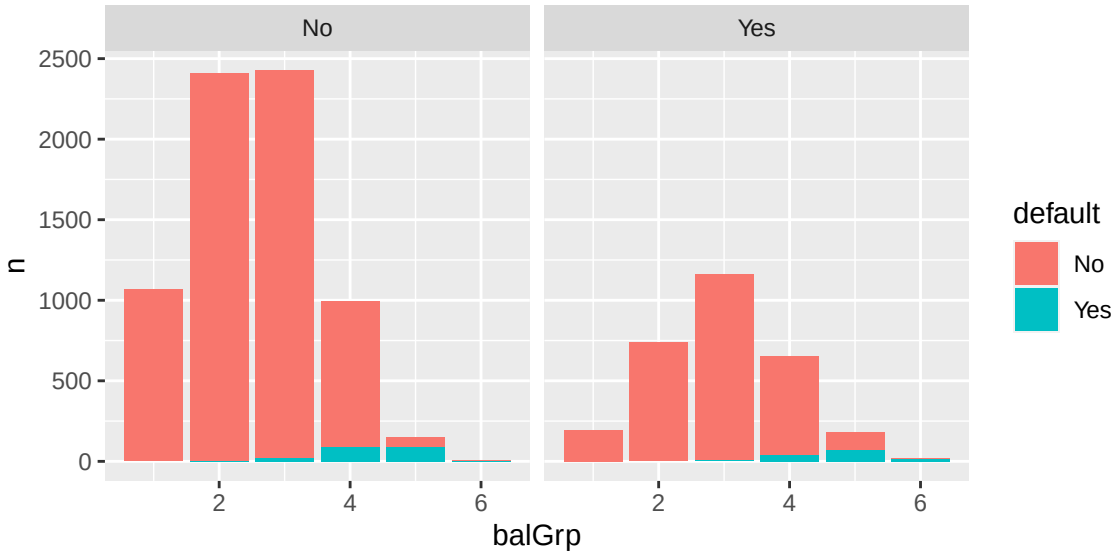


To make this easier, and to build some intuition about the data, let's break balance into discrete categories. Use the following code so that we're all on the same page:

```
dfDefault$balGrp = as.numeric(as.factor(cut_width(dfDefault$balance, 500)))

DefaultAnalysis = dfDefault %>%
  group_by(default, balGrp, student) %>%
  tally()

p = ggplot(DefaultAnalysis, aes(x = balGrp, y = n, fill = default)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ student)
p
```



Answer the following questions:

(just write your answers in the pdf)

1. $P(default = "Yes")$?
2. $P(student = "Yes")$?
3. $P(student = "Yes")$?
4. $P(default = "Yes"|student = "Yes")$
5. $P(default = "Yes"|balGrp = 3)$
6. $P(default = "Yes"|balGrp = 3, student = "Yes")$