Note: "OIS" refers to OpenIntro Statistics, the textbook for this class.

Problem 1: OIS Exercise 7.11, page 261 (9 points)

Georgianna claims that in a small city renowned for its music school, the average child takes less than 5 years of piano lessons. We have a random sample of 20 children from the city, with a mean of 4.6 years and a standard deviation of 2.2 years.

(a) Evaluate Georgianna's claim (or that the opposite might be true) using a hypothesis test (4 pts).

Prepare (2 pts)

$$H_0 : \mu \geq 5$$
$$H_A : \mu < 5$$

The sample mean, $\bar{x}$, is 4.6, and the sample standard deviation, s, is 2.2. There are 20 children in the sample. We will use a significance level of 5%.

Check (1 pt)

We have a random sample, so we can assume that the sample is independent. We cannot check normality, so we will assume the data are normally distributed and proceed with caution.

Calculate (2 pts)

```
xbar <- 4.6
s <- 2.2
n <- 20
tscore <- (xbar - 5)/(s/sqrt(n))
tscore
```

```
## [1] -0.8131156
```

```
pt(tscore, df = 19, lower.tail = T)
```

```
## [1] 0.213112
```

Interpret (2 pts)

The p-value is equal to 0.213, so we fail to reject the null hypothesis. There is not enough evidence to say that the average number of years that a child in this city takes piano lessons is less than 5.

(b) Construct a 95% confidence interval for the number of years students in this city take piano lessons, and interpret it in the context of the data (2 pts).

Note: you do not have to write down the Prepare and Check steps for (b), only Calculate and Interpret.

Calculate (1 pt)

```
tstar <- qt(0.975, df = 19)
MoE <- tstar*s/sqrt(n)
xbar - MoE
```

```
## [1] 3.570368
```

```
xbar + MoE
```

```
## [1] 5.629632
```

Interpret (1 pt)

We are 95% confident that the true mean of years of piano lessons for children living in this city is between 3.57 and 5.63.

Problem 2: OIS Exercise 6.26a, page 227 (6 points)

A news article reports that "Americans have differing views on two potentially inconvenient and invasive practices that airports could implement to uncover potential terrorist attacks." This news piece was based on a survey conducted among a random sample of 1,137 aduts nationwide, where one of the questions on the survey was "Some airports are now using 'full-body' digital x-ray machines to electronically screen passengers in airport security lines. Do you think these new x-ray machines should or should not be used at airports?" Below is a summary of responses based on party affiliation.[27]

		Party Affiliation		
		Republican	Democrat	Independent
Answer	Should	264	299	351
	Should not	38	55	77
	Don't know/No answer	16	15	22
	Total	318	369	450

(4 pts) Conduct an appropriate hypothesis test evaluating whether there is a difference in the proportion of Republicans and Democrats who think that the full-body scans should be applied in airports.[1].

Prepare (1 pt)

Let p_R be the proportion of Republicans who think full-body scans should be applied in airports, and p_D be the proportion of Democrats who think full-body scans should be applied in airports.

$$H_0 : p_R = p_D$$
$$H_A : p_R \neq p_D$$

$$\hat{p}_R = \frac{264}{318}$$
$$\hat{p}_D = \frac{299}{369}$$
$$\hat{p}_{pooled} = \frac{264 + 299}{318 + 369}$$

Check (1 pt)

We will assume that all members of the sample are independent from one another.

```
phat_pooled <- (264 + 299)/(318 + 369)
phat_pooled
```

```
## [1] 0.8195051
```

[1] S. Condon. "Poll: 4 in 5 Support Full-Body Airport Scanners". In: *CBS News* (2010)

```
phat_pooled*318
```

```
## [1] 260.6026
```

```
(1-phat_pooled)*318
```

```
## [1] 57.39738
```

```
phat_pooled*369
```

```
## [1] 302.3974
```

```
(1-phat_pooled)*369
```

```
## [1] 66.60262
```

All expected counts calculated with the pooled proportion are greater than 10, so we will proceed with the test.

Calculate (2 pts)

```
prop.test(x = c(264, 299), n = c(318, 369), correct = FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(264, 299) out of c(318, 369)
## X-squared = 0.45685, df = 1, p-value = 0.4991
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.03758321  0.07736436
## sample estimates:
##    prop 1    prop 2
## 0.8301887 0.8102981
```

Interpret (2 pts)

Since the p-value is greater than 0.10, we fail to reject the null hypothesis. There is not enough evidence to say the proportions of Republicans and Democrats who think full-body scans should be applied in airports are different.

Problem 3 (10 points)

Prices of diamonds are determined by what is known as the 4 C's: cut, clarity, color, and carat weight. The prices of diamonds of up as the carat weight increases, but the increase is not smooth. For example, the difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but the price of a 1 carat diamond tends to be much higher than the price of a 0.99 carat diamond.

In this question, we use two random samples of diamonds with carat weights 0.99 and 1. In order to be able to compare equivalent units, we calculate the price per carat, by dividing the price for each diamond by 100 times its weight in carats. That is, for a 0.99 carat diamond, we divide the price by 0.99. For a 1 carat diamond, we divide the price by 100. The variable to analyze is stored in the variable `ppc`.

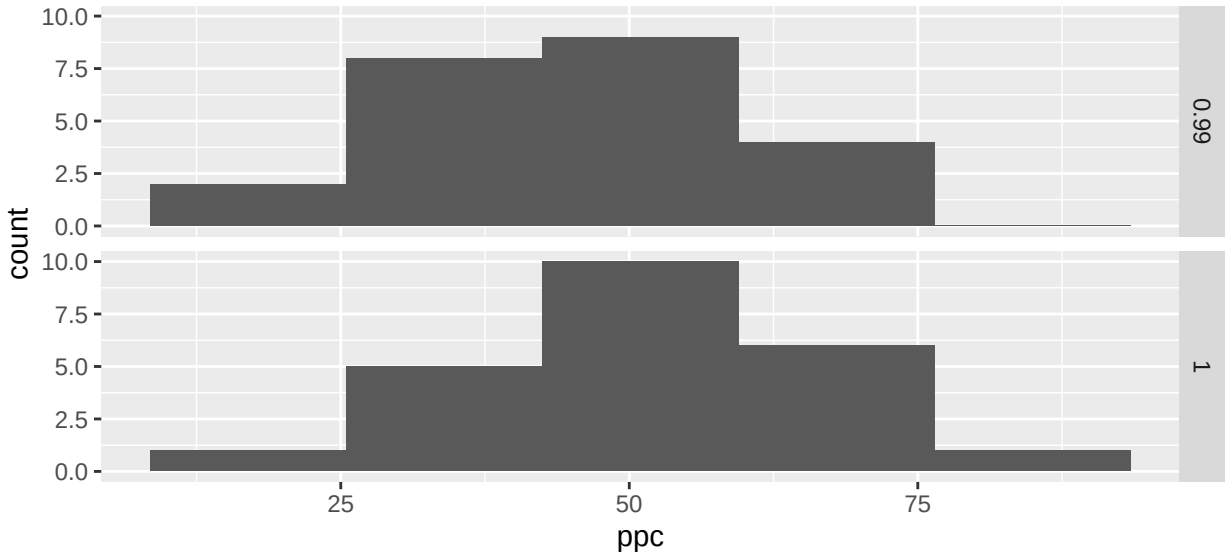
```
diamonds <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/diamonds.csv")

diamonds <- diamonds[c(386, 424, 1276, 3171, 4308, 6025, 6168, 6217, 6226, 6460, 6558,
                       6955, 7006, 7816, 8482, 9022, 9360, 10225, 10309, 10416, 10540,
                       10752, 10805, 11090, 11357, 11884, 12018, 12602, 13931, 13940,
                       14125, 14559, 14608, 14777, 14991, 15159, 15170, 15230, 15907,
                       16294, 16550, 16899, 20147, 46680, 47803, 48260),]

diamonds$ppc <- as.numeric(diamonds$price/(100*diamonds$carat))
```

(a) Create two histograms of the price per carat, one for the group of 0.99 carat diamonds and one for the group of 1 carat diamonds. The histograms should have five bins (4 pts).

Hint: You can either create two new data frames using the `subset()` command or use the command `facet_grid()` to automatically split the data with `ggplot`.

(b) Comment on the two histogram (2 pts).

```
ggplot(data = diamonds, aes(x = ppc)) +
  facet_grid(carat~.) +
  geom_histogram(bins = 5)
```



Both histograms are centered around \$50, and range from \$0 to \$100. They are approximately normal with no outlying values.

(c) Conduct an appropriate hypothesis test evaluating whether there is a difference in price per carat for 0.99 carat diamonds and 1 carat diamonds. Use your results from parts (a) and (b) to check the conditions (4 pts).

Prepare (1 pt)

Let $\mu_{0.99}$ be the mean price per carat for 0.99 carat diamonds μ_1 be the mean price per carat for 1 carat diamonds.

$$H_0 : \mu_{0.99} = \mu_1$$
$$H_A : \mu_{0.99} \neq \mu_1$$

$$\bar{x}_{0.99} = 44.5,\ s_{0.99} = 13.3,\ n_{0.99} = 23$$

$$\bar{x}_1 = 51.4,\ s_1 = 13.7,\ n_1 = 23$$

We will use a significance level of $\alpha = 0.05$.

Check (1 pt)

Because this is a random sample, all observations are independent within and between groups. In addition, the histograms from part (a) look normal, so all conditions for the test are met.

Calculate (1 pt)

```
t.test(data = diamonds, ppc ~ carat)
```

```
##
##  Welch Two Sample t-test
##
## data:  ppc by carat
## t = -1.7345, df = 43.97, p-value = 0.08984
## alternative hypothesis: true difference in means between group 0.99 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -14.930004   1.118401
## sample estimates:
## mean in group 0.99    mean in group 1
##           44.50681           51.41261
```

Interpret (1 pt)

The test statistic is -1.7345 and the p-value is 0.08984. Since the p-value is greater than the significance level, we fail to reject the null hypothesis. There is not enough evidence to say that the mean price per carat is different for 0.99 carat diamonds and 1 carat diamonds.

This problem was inspired by OIS Exercise 7.24, page 274.