R Lab 3: Chi-square Goodness of Fit Test in R

Is the census report accurate?

Team Captain: NAME HERE Facilitator: NAME HERE
Recorder: NAME HERE Resource Manager: NAME HERE

Invalid Date

0. **Load the `infer` package below `library(tidyverse)`**

```
# Load in the packages necessary for our analysis: tidyverse, infer
library(tidyverse)
```

The accuracy of a census report for a city in southern California was questioned by some government officials. A random sample of 1,215 people living in the city was used to check the census report. The census results (expected probabilities) are given below:

	Black	Asian	Anglo	Latino/Latina	Indigenous	All others
Census Result	10%	3%	38%	41%	6%	2%

Research question: Do the counts obtained via the random sample call into question the accuracy of the census report?

1. **Read in the data and call it `ca_sample`. Additionally, print the top 6 rows to inspect your data.**

```
________ <- read_csv("data/________.csv")
```

```
Error: <text>:1:2: unexpected input
1: __
     ^
```

2. **What is the observational unit for this study?**

   Answer goes here.

3. **Create a summary table of the number of individuals in the sample who self-identified with each ethnic `origin`.**

4. **Using the skills you learned class last week, create a bar plot of the observed *proportions* of individuals in our sample who self-identified with each ethnic `origin`.**

5. **Notice the ordering of origin categories. Why do you think "All others" is first and "Latino/Latina" is last in our list?**

   Answer goes here.

6. **As a statistician, Lily questions the "All others" category. What might be some data and ethical concerns surrounding reporting and aggregating data in this way?**

   Answer goes here.

7. **Write the null and alternative hypotheses for investigating this research question.**

   Null:

   Alternative:

8. **How many of the 1,215 people do we expect to see of each ethnic origin if the results of the census are accurate?**

	Black	Asian	Anglo	Latino/Latina	Indigenous	All others
Expected Count						

9. **Check the conditions for using the Chi-square distribution to test our hypotheses.**

   Answer goes here.

10. **Carry out the Chi-square Goodness of Fit Test and find the p-value for investigating this research question.**

```
chisq_test(x = ________,
           response = _________,
           p = c("________" = ____,
                 "________" = ____,
                 "________" = ____,
                 "________" = ____,
```

```
                  "--------" = ____,
                  "--------" = ____,
            )

 )
```

```
Error: <text>:1:17: unexpected input
1: chisq_test(x = __
                    ^
```

10. **From the output, identify the observed test statistic, degrees of freedom, and p-value.**

    Answer goes here.

11. **Write a conclusion in the context of the research question.**

    Answer goes here.