STA130H1S – Fall 2022

Problem Set 7

() and STA130 Professors

Instructions

Complete the exercises in this .Rmd file and submit your .Rmd and .pdf output through Quercus on Thursday, November 3rd by 5:00 p.m. ET.

```
library(tidyverse)
```

Part 1: Broadway

Question 1: Broadway

Lin-Manuel Miranda was nominated for "Best Original Song" for the March 27, 2022 the Academy Awards (also known as the Oscars) for his work on the Disney movie Encanto. Miranda had already won an Emmy, Grammy, and Tony (mostly for his work on the broadway musical "Hamilton"), so he was very close to the (EGOT)[https://www.vanityfair.com/hollywood/2022/02/oscar-nominations-2022-will-lin-manuel-miranda-finally-egot-for-encanto] (Emmy, Grammy, Oscar and Tony), a rare occurrence as only 16 people have won all four awards see here. Unfortunately, Miranda did not win the Oscar in 2022. Perhaps he will soon!

In this question we will look at a sample of weekly broadway musical data available in the `broadway.csv`. This data set contains a sample of Broadway musical information for 500 weeks from 1985 to 2020. In this data set an observation is one broadway musical in a particular week (ending on a Sunday). Variables of interest are:

- show: Name of the broadway musical/show.
- Hamilton: indicates whether the musical is "Hamilton" or not.
- week_ending: Date of the end of the weekly measurement period. Always a Sunday.
- weekly_gross_overall: Weekly box office gross for all shows.
- avg_ticket_price: Average price of tickets sold in a particular week.
- top_ticket_price: Highest price of tickets sold in a particular week.
- seats_sold: Total seats sold for all performances and previews in a particular week.
- pct_capacity: Percent of theatre capacity sold. Shows can exceed 100% capacity by selling standing room tickets.

Let's explore different ways to estimate the average ticket price for Broadway shows!

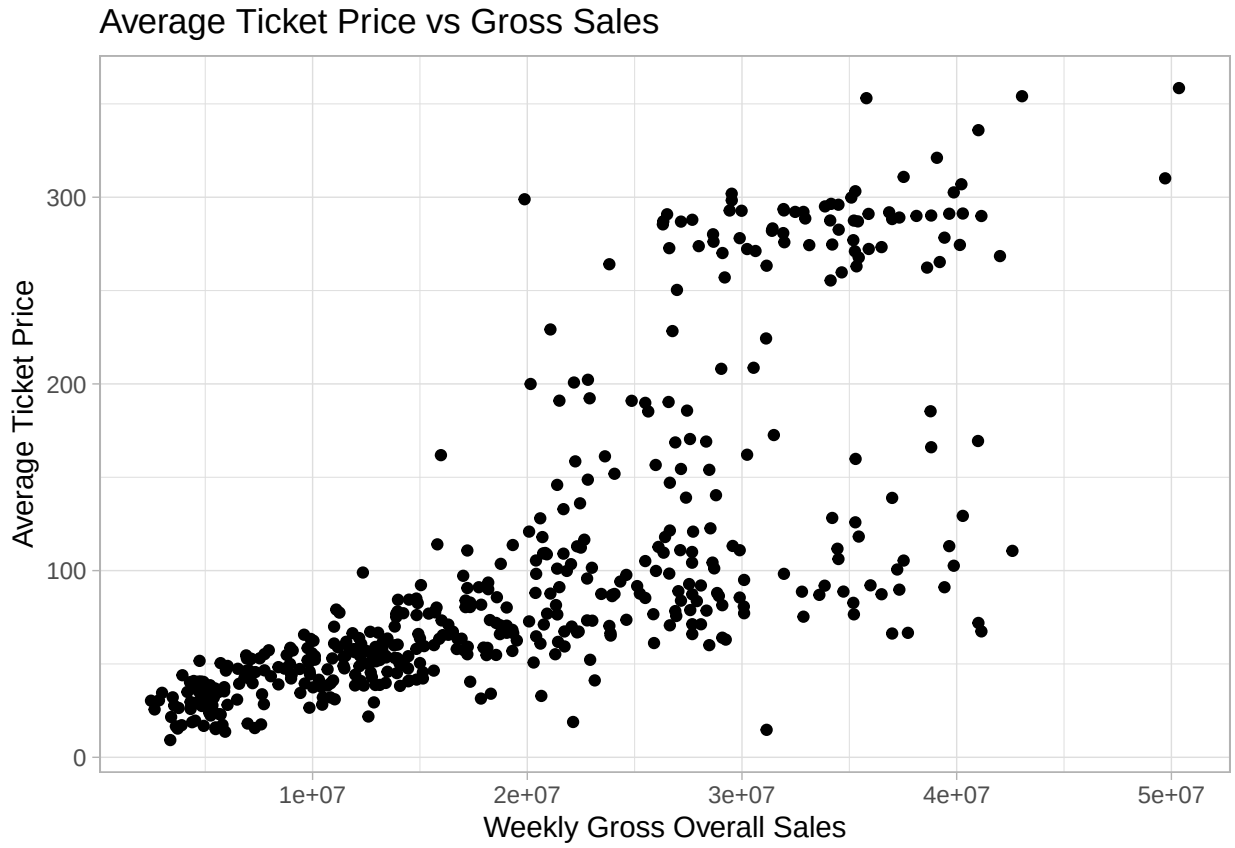
```
broadway_data <- read_csv("broadway.csv")
glimpse(broadway_data)
```

```
## Rows: 500
## Columns: 8
## $ show                <chr> "La Cage aux Folles", "42nd Street", "42nd Street~
## $ Hamilton            <chr> "No", "No", "No", "No", "No", "No", "No", "No", "~
## $ week_ending         <date> 1985-07-28, 1985-09-08, 1985-09-15, 1985-12-15, ~
```

```
## $ weekly_gross_overall <dbl> 2989271, 2474396, 2844860, 4169643, 3555363, 3632~
## $ avg_ticket_price     <dbl> 34.54, 30.31, 30.50, 35.00, 27.74, 16.60, 17.19, ~
## $ top_ticket_price     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ seats_sold           <dbl> 11841, 7251, 7890, 10846, 2803, 2204, 5740, 10861~
## $ pct_capacity         <dbl> 0.8795, 0.5477, 0.5959, 0.8056, 0.2967, 0.4364, 0~
```

(a) Make a plot showing the relationship between the average ticket price (on the y-axis) and the weekly gross overall sales (on the x-axis).

```
# code you answer here
broadway_data%>%
ggplot(aes(x=weekly_gross_overall,y=avg_ticket_price)) +
geom_point() + theme_light() +
labs(title="Average Ticket Price vs Gross Sales",
x="Weekly Gross Overall Sales",
y="Average Ticket Price")
```



(b) Briefly explain whether or not you think it is appropriate to characterize and summarize the association in the above plot with a straight line.

We can say that the above plot can be summarized using the above plot because we can see a general trend being an upward slopping straight line.

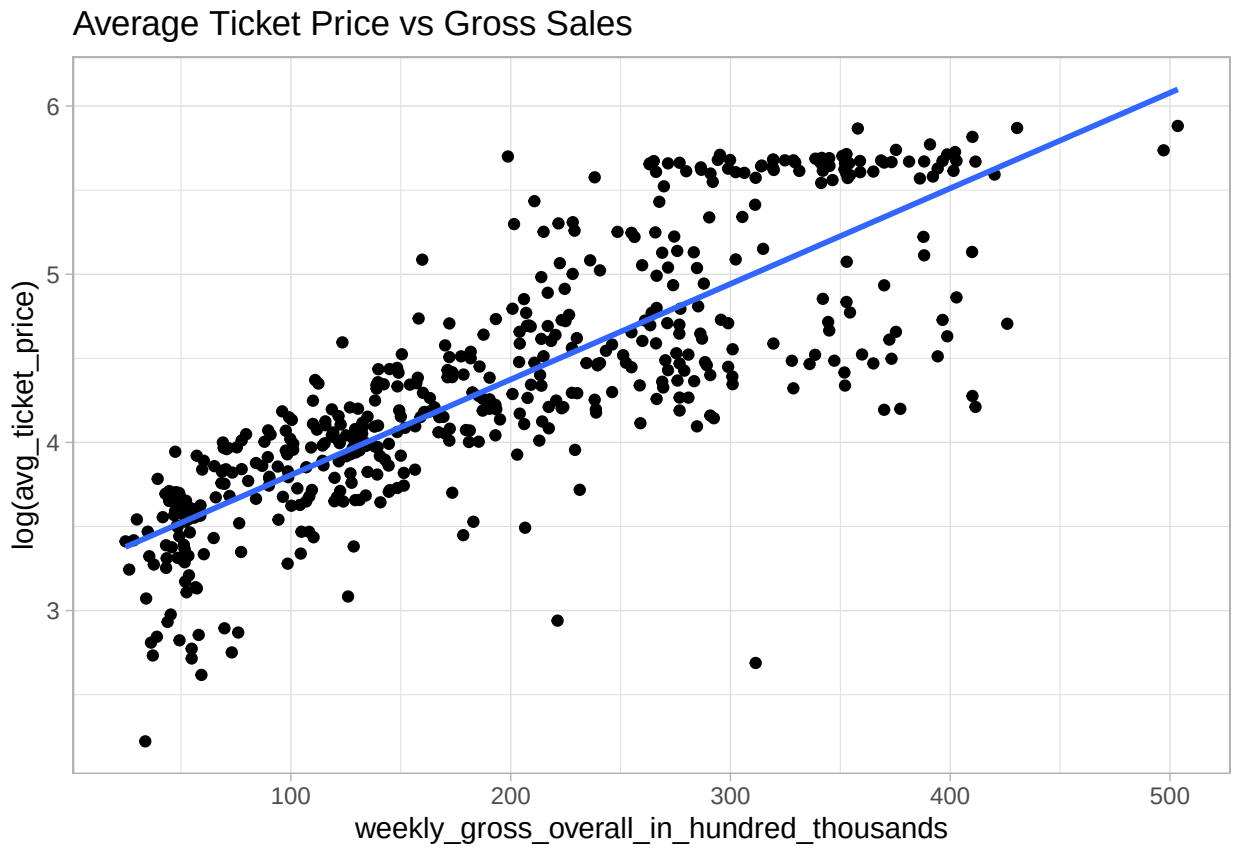
(c) Use the `mutate()` function to add the new variables `log_avg_ticket_price = log(weekly_gross_overall)` and `weekly_gross_overall_in_hundred_thousands=weekly_gross_overall/100000` to the dataset.

```
broadway_data %>%
  mutate(log_avg_ticket_price = log(avg_ticket_price),weekly_gross_overall_in_hundred_thousands=weekly_g
```

(d) Plot the association between `log_avg_ticket_price` (on the y-axis) and `weekly_gross_overall_in_hundred_th` (on the x-axis) and use `geom_smooth(method=lm, se=FALSE)` to add a line of best fit to the plot. Describe the association you observed in the plot.

Note: You will learn more about transforming variables in future courses and are not required to be able to explain why we've done this here. You can just treat `log_avg_ticket_price` as we have other variables in class and refer to it as "the natural log of average ticket price".

```
broadway%>%
ggplot(aes(x=weekly_gross_overall_in_hundred_thousands,y=log_avg_ticket_price)) +
geom_point() + theme_light() +geom_smooth(method=lm, se=FALSE)+
labs(title="Average Ticket Price vs Gross Sales",
x="weekly_gross_overall_in_hundred_thousands",
y="log(avg_ticket_price)")
```



I have observed a positive correlation which indicates that most of the data conforms to the trend . We can also confirm that line may not represent the points near the top for they are far appart. However, this will be a strong correlation. only till x=300

(e) Use the `cor()` function to calculate the correlation between `broadway_data$log_avg_ticket_price` and `broadway_data$weekly_gross_overall_in_hundred_thousands`.

```
# code you answer here
cor(broadway$log_avg_ticket_price,broadway$weekly_gross_overall_in_hundred_thousands)
```

```
## [1] 0.8154224
```

(f) Write down a simple linear regression model specification with response `log(avg_ticket_price)` and explanatory varaible `weekly_gross_overall_in_hundred_thousands`. Explain each compnwnt of the model.

$$\hat{y} = \hat{\beta}_0 + x_i\hat{\beta}_1 + \hat{\epsilon}_i$$

y with hat is the outcome from the equation. B_0 is the y-intercept. xi is the indepentant values ussually observations. B_1 is the rate which indicates correlation or no correlation. Epsilon indicates the noise in terms of how data points varry from the graph

Hint: If you copy math equations from another software into your .Rmd document, you'll get errors when trying to knit. Instead, you should type your math equations directly in your .Rmd document. Here are some tips and examples for doing this:

1. In a .Rmd document, math equations and symbols must be typed between dollar symbols ($).
2. If you want your equation/symbol to appear in the middle of a sentence, use only one dollar sign before and one dollar sign after. For example, we can typeset beta-hat-0 in .Rmd as $\hat{\beta}_0$.
3. If you want your equation to appear on a line on its own, type it on a separate line and put two dollar signs at the begining and the end. For example,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1$$

4. A few other useful symbols you may need in this question are epsilon (ϵ), "not equal" ($\neq$), and superscripts (e.g. i^{th}).

(g) State the null and alternative hypotheses you would use assess whether the slope of the linear regression model where weekly gross overall income in 1000s is predicting the log average ticket price.

$$\hat{\beta}_1 = 0$$
$$\hat{\beta}_1 \neq 0$$

(h) Use R to fit the linear model that corresponds with your line of best fit above. Report the fitted equation of the line. Interpret the regression coefficients in the context of this data AND make a conclusion about the hypotheses you defined above.

As stated by the lm function ouputs below the estimate for the estimate we can based on that conclude that equation is y=0.006x+3.238 (rounded to 3 decimal). The regression coeffiencets state that the intercept meaaning the point at which y=0 is 3.238 and that a rise/run of the line is 0.006. This indicates to us the B_0 is 3.238 and B_1 is 0.0006. This means that our alternate hypothesis in this instance is that B_1=0.0006

```
least_squares_fit <- lm( log_avg_ticket_price ~ weekly_gross_overall_in_hundred_thousands , data=broadwa
summary(least_squares_fit)$coefficients
```

```
##                                          Estimate   Std. Error  t value
## (Intercept)                             3.237767214 0.0404308708 80.08156
## weekly_gross_overall_in_hundred_thousands 0.005682641 0.0001807721 31.43539
```

```
##                                                 Pr(>|t|)
## (Intercept)                                 1.357759e-286
## weekly_gross_overall_in_hundred_thousands 2.542862e-120
```

(i) Get the R^2 for your model and write one sentence interpreting it in context.

R^2 determines the strenght of the model and the varriance of the dependant variables. A low R^2 indicates that the line does not represent the graph well.

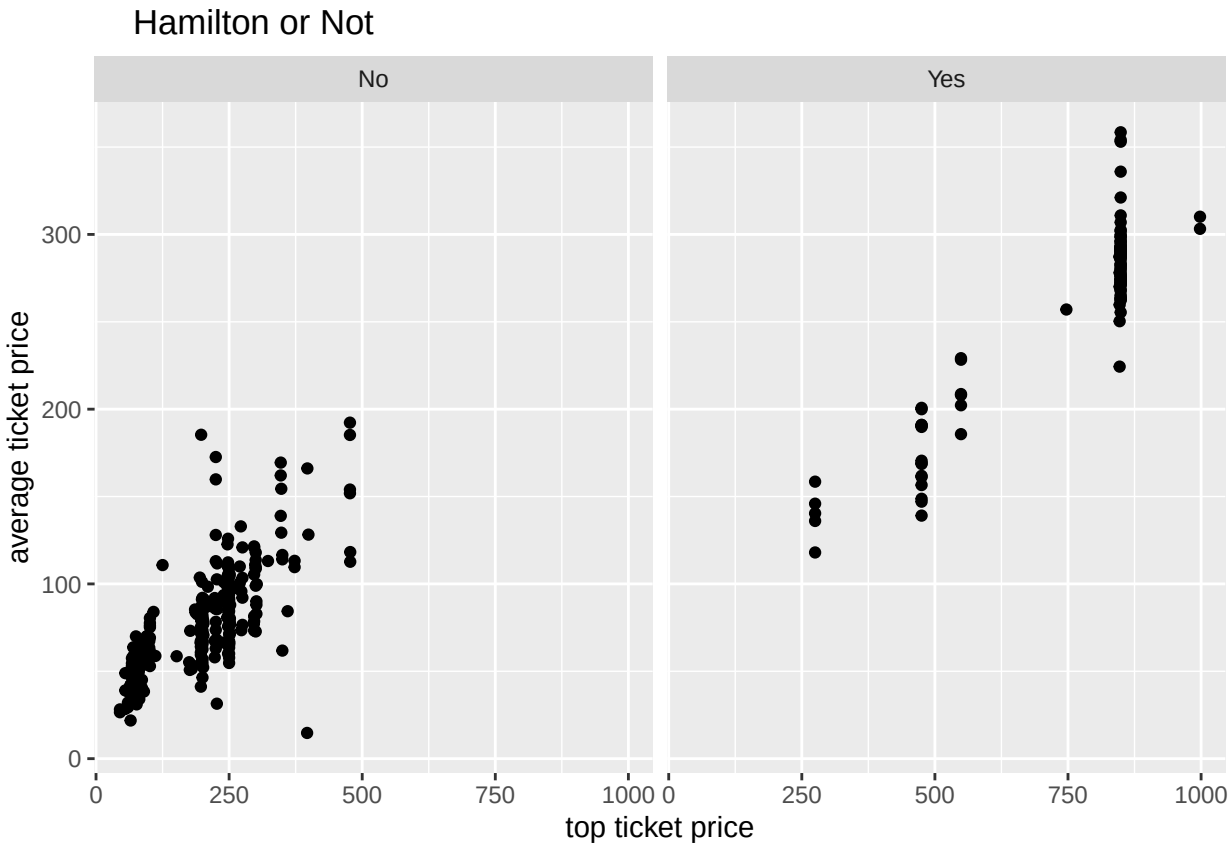
```
# code you answer here
least_squares_fit <- lm( log_avg_ticket_price ~ weekly_gross_overall_in_hundred_thousands , data=broadwa
1-as.numeric(summary(least_squares_fit)$r.squared)
```

```
## [1] 0.3350864
```

(j) Create a plot of the association between `top_ticket_price` (on the y-axis) and `top_ticket_price` (on the x-axis) faceted by whether the musical was "Hamilton" or not using `facet_wrap(~Hamilton)`.

Comment about question- x and y axis are the same. TA instructed to use average ticket price

```
# code you answer here
 broadway%>%
ggplot(aes(x=top_ticket_price, y=avg_ticket_price)) +
geom_point() + facet_wrap(~Hamilton) +
labs(title="    Hamilton or Not    ",x="top ticket price",
y="average ticket price")
```



(k) Calculate the correlation between top ticket price and average ticket price for both Hamilton and non-Hamilton musicals using the `group_by`, `summarise()` and `cor(top_ticket_price, top_ticket_price, na.rm=TRUE)`.

```
# code you answer here
broadway %>% group_by(Hamilton) %>% summarise(correlation=cor(top_ticket_price, avg_ticket_price),na.rm=
```

```
## # A tibble: 2 x 2
##   Hamilton correlation
##   <chr>          <dbl>
## 1 No            NA
## 2 Yes            0.922
```

Part 2: Optional Indicator Variable Simple Linear Regression

You may complete these questions for practice if you wish. ***You are not required to complete these questions as they ARE NOT included as part of your mark.***

Note that there is a Part 3 below which is "less" optional.

Question 2: Starbucks

The `starbucks.csv` dataset contains data on calories and carbohydrates (in grams) in Starbucks food menu items.

```
starbucksdata<-read_csv("starbucks.csv")
glimpse(starbucksdata)
```

```
## Rows: 77
## Columns: 7
## $ item     <chr> "8-Grain Roll", "Apple Bran Muffin", "Apple Fritter", "Banana~
## $ calories <dbl> 350, 350, 420, 490, 130, 370, 460, 370, 310, 420, 380, 320, 3~
## $ fat      <dbl> 8, 9, 20, 19, 6, 14, 22, 14, 18, 25, 17, 12, 17, 21, 5, 18, 1~
## $ carb     <dbl> 67, 64, 59, 75, 17, 47, 61, 55, 32, 39, 51, 53, 34, 57, 52, 7~
## $ fiber    <dbl> 5, 7, 0, 4, 0, 5, 2, 0, 0, 0, 2, 3, 2, 2, 3, 3, 2, 3, 0, 2, 0~
## $ protein  <dbl> 10, 6, 5, 7, 0, 6, 7, 6, 5, 7, 4, 6, 5, 5, 12, 7, 8, 6, 0, 10~
## $ type     <chr> "bakery", "bakery", "bakery", "bakery", "bakery", "bakery", "~
```

(a) Produce a plot that shows the association between carbohydrates and calories in Starbucks menu items. Describe this association.

REPLACE THIS TEXT WITH YOUR ANSWER

```
# code you answer here
```

(b) Before calculating anything, estimate the correlation coefficient between carbohydrates and calorie content in Starbucks menu items based on the plot you produced in (a). Justify your answer.

REPLACE THIS TEXT WITH YOUR ANSWER

(c) Calculate the correlation between carbohydrate and calorie content of Starbucks menu items. How does this compare to your estimate in part (b)?

REPLACE THIS TEXT WITH YOUR ANSWER

```
# code you answer here
```

(d) Write down a simple linear regression model specification for the content of Starbucks menu items with `calories` as tbe response variable and `carb` as the explanatory variable. Explain each term in the model.

REPLACE THIS TEXT WITH YOUR ANSWER

(e) Describe what the test $H_0 : \beta_1 = 0$ vs $H_A : \beta_1 \neq 0$ is testing.

REPLACE THIS TEXT WITH YOUR ANSWER

(f) Use R to fit the regression model in (d) to these data. Report the fitted regression line. Interpret the regression coefficients in the context of this study AND make a conclusion about the hypotheses you defined above.

REPLACE THIS TEXT WITH YOUR ANSWER

```
# code you answer here
```

(g) Add the estimated linear regression line that you calculated in (f) to the plot you generated in (a). Compute the coefficient of determination, R^2. How well does the linear regression line seem to capture the relationship between `carb` and `calories`? Justify your answer.

REPLACE THIS TEXT WITH YOUR ANSWER

```
# code you answer here
```

(h) Based on the Starbucks data, create a new dataset called `starbucks_lunch` which only contains food items which are only of the "sandwich" or "bistro box" type. Create a boxplot comparing the distribution of calories for these two types of items.

```
# code you answer here
```

(i) Fit a linear regression model to test whether there is a difference in mean calories for items of type "bistro box" and items of type "sandwich". Write a sentence summarizing your conclusion.

REPLACE THIS TEXT WITH YOUR ANSWER

```
# code you answer here
```

Part 3: Preparation for Week 10 Tutorial [Get Started Early!!!]

Part 3 is VERY BIG with many, MANY questions. These questions WILL NOT be graded as part of your Problem Set 7 submission (in Week 8); rather, they will be graded as part of your Problem Set 8 submission (in Week 10). However, because there are SO MANY questions in this sequence of exercises, is it HIGHLY RECOMMENDED that you complete as many of these questions as you can now so that you will have all the answers ready in time for the Week 10 submission. If you do not complete these questions sooner rather than later you will put your ability to complete them in jeopardy.

- ***Many thanks to head TA Quin Xie for inspiring this series of questions through her choice of tutorial reading and presentation paper, and for providing feedback on initial drafts of these questions which led to substantial revision of the development of these questions.***

In the week 10 tutorial, you will (in groups) present the analysis of Ross *et al.* (Nature 2022) "Women are Credited Less in Science than are Men", available on quercus and online through the UT library. In preparation for that, we'll explore the story of Rosalind Franklin which is noted by Ross *et al.* to frame their work.

The discovery, or proposal, of the double-helix structure of DNA was published in 1953. In 1962 the Nobel Prize for Physiology or Medicine "for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material" was awarded to three men: Wilkins, Crick, and Watson. A woman, Rosalind Franklin was not included in the award. However, Watson himself donated a commemorative statute upon which the base (i.e., foundation) read "The double helix model was supported by the work of Rosalind Franklin and Maurice Wilkins" while the helices themselves contain the statement "The structure of DNA was discovered in 1953 by Francis Crick and James Watson while Watson lived here at Clare [College, Cambridge, England]". The remainder of the inscriptions on the statue are "The molecule of DNA has two helical strands that are linked by base pairs Adenine – Thymine or Guanine – Cytosine" and "These strands unravel during cell reproduction. Genes are encoded in the sequence of bases". Sadly, there is a technical reason for why Rosalind Franklin is not a co-recipient on the award: the Nobel Prize cannot be awarded posthumously, and Rosalind Franklin died in 1958 (at only 37).

Question 2: The Double-Helix and Rosalind Franklin

The seminal 1953 manuscript presenting the double-helix structure of DNA was authored by Watson and Crick. Wilkins, who was not an author on the manuscript, was given access to the the manuscript to review and his feedback was incorporated into the final manuscript. Rosalind Franklin was also not an author on the manuscript, and/but was not asked to review or provide feedback on the manuscript during the drafting stage.

The Ross *et al.* manuscript states "... [Rosalind Franklin] was wrongfully denied authorship on the original Crick and Watson [1953] paper."

The following statements from Watkins and Crick shed light on the nature of the above claim.

- Watson: "The truth is that we should have got the structure in the fall of '51 (rather than 53). There was enough data. We wouldn't have been able to say with finality that it was right because that came with Rosalind's X-ray work: that was the proof it was right."
- Watson: "[W]hen Rosy's [Rosalind Franklin's] talk on DNA rolled about, I had learned enough crystallographic argument to follow much of her lecture. Most important, I knew what to focus attention upon. Six weeks of listening to Francis [Crick] had made me realize that the crux of the matter was whether Rosy's new X-ray pictures would lend any support for a helical DNA structure. The really relevant experimental details were those which might provide clues in constructing molecular models. It took, however, only a few minutes of listening to Rosy to realize that her determined mind had set upon a different course of action. ... It was downright obvious to her that the only way to establish the DNA structure was by pure crystallographic approaches. As model building did not appeal to her... The idea of using tinker-toy-like models to solve biological structures was clearly a last resort. Of course Rosy knew of Linus' [Pauling's] success [based on model building] but saw no obvious reason to ape his mannerisms."
- Watson: "I was shown Rosalind Franklin's x-ray photograph and, Whooo! that was a helix, and a month later we had the structure, and Wilkins should never have shown me the thing... I didn't go into the drawer and steal it. It was shown to me ... the Franklin photograph was the key event ... psychologically, it mobilised us back into action... "
- Watson: "Rosy, of course, did not directly give us her data. For that matter, no one at Kings realized they were in our hands."
- Crick: "We are most heavily indebted in this respect to the King's College Group [where both Wilkins and Rosalind Franklin worked independently], and we wish to point out that without this data the formulation of our structure would have been most unlikely, if not impossible."
- Crick: "Our belief is that she [Rosalind Franklin] didn't realize until the structure came out how important DNA was. For her, it was just another problem."

- Crick (as reported by New York Times science writer Nicholas Wade): "I asked [Crick] once if he thought he had come into Rosalind Franklin's information in a fair way and he said yes, he thought it was fair" and "Crick believed it was legitimate to use data that had been publicly presented."

For her part, Rosalind Franklin did not believe that DNA had a helix structure, writing in a subsequent paper that "discrepancies prevent us from accepting it in detail"; and, previously (along with her PhD student Raymond Gosling) penning and signing a satirical obituary rejecting the double-helix hypothesis:

> "It is with great regret that we have to announce the death, on Friday 18th July 1952 of DNA helix (crystalline). Death followed a protracted illness which an intensive course of Besselised injections had failed to relieve. A memorial service will be held next Monday or Tuesday. It is hoped that Dr M H F Wilkins will speak in memory of the late helix."

(a) Explore your initial reaction to the Ross *et al.* statement "...[Rosalind Franklin] was wrongfully denied authorship on the original Crick and Watson paper." What interpretation would justify this statement? What interpretation would disagree with this statement?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Which interpretation do you favor and why? Do you think authorship should evaluated on the basis of conceptualization of the double-helix? Or based on generating the confirmatory and catalyzing data as well?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 3: Personal Bias

STA130 professor Scott Schwartz, upon reading the abstract of the Ross *et al.* manuscript, specifically

> "[W]omen in research teams are significantly less likely than men to be credited with authorship... women are significantly less likely to be named on a given article [13.24% gap] or patent [58.40% gap] produced by their team relative to their male peers. The gender gap in attribution is present across most scientific fields and almost all career stages."

had the following to say:

> "I don't see how rightful authorship to the level reported by this paper could be possible... statistics and math papers do not have doubt as to who authors should be: authorship is not about contribution of ideas, but contributions to manuscript writing (with contributing ideas cited and referenced). I do see possible authorship bias happening from a mechanism whereby women are not invited to collaborate on a paper when they should be though."

(a) What biases does the statement appear to have? Does it rush to judgement quickly, or does it appear to read the abstract honestly and charitably? [Your grade will not be negatively affected if you think Prof Schwartz's statement problematic and provide criticisms of it.]

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Was Professor Schwartz statement compelling to any degree? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) Can you identify any of your own personal biases and principles that lead to your answers to the above questions?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 4: The "Appeal to Authority" Fallacy

In the previous question Professor Schwartz stated his opinion that "authorship is not about contribution of ideas, but contributions to manuscript writing (with contributing ideas cited and referenced)". Explaining this further, he states

> "I have been involved in scientific research since 2005 when I started my PhD program, including more than 5 years embedded in hard science wetlabs doing interdisciplinary collaborative research. You often see unpublished work through presentations and it informs your thinking. But this doesn't generally mean you'll invite the person to be a co-author on a manuscript. You only do that if you're planning to actually collaborate and work together. If you end up inspired or relying on someone elses ideas, you are (morally) obligated to acknowledge those and recognize them appropriately. Certainly you could explicitly ask people if you can reference them, and ask them for feedback in the manner you are representing and acknowledging them, but it's not mandatory."

In the case of Rosalind Franklin, Professor Schwartz says (based on the on the actual references and acknowledgments of Rosalind Franklin by Watson and Crick in the 1953 manuscript, which you will see in Question 7 below) that he doesn't agree with Ross *et al.* that "... [Rosalind Franklin] was wrongfully denied authorship on the original Crick and Watson paper."

(a) Do you view Professor Schwartz as an authority figure with an especially relevant and well-informed opinion on this topic? Or do you think he's not an authority figure in this context? [Your grade will not be negatively affected if you are critical about Prof Schwartz or his opinions.]

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Relative to Professor Schwartz's opinion, how important, or in what ways might it be important whether or not Professor Schwartz is an authority figure on this topic?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) To what degree does your opinion about Professor Schwartz's statement depend on your evaluation of whether or not he is a legitimate well-informed authority figure on this topic?

REPLACE THIS TEXT WITH YOUR ANSWER

(d) To what degree do you think Ross *et al.* are authority figures relative to their claim? And to what degree might your opinion (and the statement "wrongfully denied authorship" from the manuscript) be influenced by this or other forms of bias you might have?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 5: Representational Motivation (Bias)

In 1968 (after the initial 1953 manuscript and 1962 Nobel Prize) Watson published a memoir called The Double Helix. Shortly thereafter the article "Rosalind Franklin and the discovery of the structure of DNA" by Aaron Klug (Nature 1968) provided a fuller discussion of "Dr Franklin's contribution to the discovery of the structure of DNA in the light of accounts given by Professor Watson in his book The Double Helix and by Dr Hamilton in a recent article in Nature... [which] does not do justice to Franklin's work". Both Watson's memoir and Klug's paper brought attention and awareness to the late Rosalind Franklin and began to establish public recognition and acknowledgment for her work. Then, in 1975, Rosalind Franklin's friend Ann Sayre wrote the biography "Rosalind Franklin and DNA" which critiqued the (often blatantly) chauvinistic characterizations of Watson's book and raised general issues of sexism related to Rosalind Franklin's experience, giving rise to the *Wronged Heroine* interpretation of Rosalind Franklin that is commonly encountered today. Current commentators such as Lynne Osman Elkin endorse this perspective, while others such as Nicholas Wade are

dismissive of it (while admitting that "The acknowledgment in [the 1953 Crick and Watson] paper ... was barely adequate"). Still others, such as Brenda Maddox, take a more nuanced perspective feeling that the *Wronged Heroine* narrative continues to undermine the totality of scientific achievements and accomplishments Rosalind Franklin was able to make in her short, but amazingly prolific lifetime.

None of the above scholars, however, have publicly endorsed a statement regarding the authorship on the 1953 paper such as that made by Ross *et al.* that "...[Rosalind Franklin] was wrongfully denied authorship on the original Crick and Watson paper." Instead, their public statements are consistent (to greater and lesser degrees) with the first part of the Ross *et al.* sentence which says "Franklin's pivotal contribution to the discovery of the structure of DNA initially went unrecognized..." (and references the Klug manuscript, apparently referring to the statement "The importance of Franklin's work has been lost sight of, partly because of her untimely death"). Indeed, even the hypothetical scenario where Rosalind Franklin is an author on the original manuscript is not necessarily even sufficient to address the concern that her contributions are not appropriately acknowledged. For example, Wilkins was awarded the Nobel Prize, but the general public understanding is that the discovery was that of "Watson and Crick", not "and Wilkins". So a hypothetical analogous treatment of Rosalind Franklin as Wilkens could simply have still have not proven to insufficient appropriate acknowledgment.

(a) Given the general consensus among Rosalind Franklin scholars that the issue is a question of recognition as opposed to authorship, why do you think it is that Ross *et al.* chose to frame their statement in terms of authorship as opposed to recognition?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) What is the difference between misrepresenting something and giving personal opinions about a topic? Is there a way to caution when the intention is to provide an opinion?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) What is the difference between misrepresenting something you are reporting, versus misrepresenting what someone else has reported, if anything?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 6: Hedging

Previously you've watched this 7-minute video introduction to hedging. But **now watch it again** *with a new perspective in mind.* Of course, as previously noted the last time this video was introduced

> "Hedging is helpful whenever you can't say something is 100% one way or another, as is often the case. In statistics, hedging should always be used with respect to the limitations of data and the strength and generalizability of the conclusions."

but also, hedging is a way to overcome hesitation and defensiveness people can experience when confronted with opinions that challenge their personal biases and *a priori* beliefs. Approaching conversations with more open-minded language can keep people from "rushing to judgment" and "shutting down" and more willing to engage and listen.

(a) Do you think any of the statements you've encountered so far in the preceding questions in Part 3 could have benefited from being made more approachable and palatable by using hedging? If so, which ones and how so?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 7: Plagiarism

Previously you've watched this 8-minute video introduction to plagiarism. Plagiarism is an extremely important topic for academic integrity, and citing the reference material you rely on makes your own work come across even more convincingly. So, **watch this video again**, *and try to pick up even more than what you remember learning last time you watched it.* For very important topics, like plagiarism, returning to review the topic is a very important tool for deeply integrating the topic into your own thinking and actions.

Let us now return to Rosalind Franklin, and consider the notion of plagiarism in her context.

As noted in a commentary,

> "Alongside the Watson-Crick paper in the April 25, 1953, issue of Nature were separately published papers by scientists Maurice Wilkins and Rosalind Franklin of King's College, who worked independently of each other. The Wilkins and Franklin papers described the X-ray crystallography evidence that helped Watson and Crick devise their structure. The authors of the three papers, their lab chiefs, and the editors of Nature agreed that all three would be published in the same issue."

and New York Times science reporter Nicholas Wade, in a panel discussion with preeminent Rosalind Franklin scholar Lynne Osman Elkin, agreed, saying:

> "There were three papers published in Nature, the first by Watson and Crick. The second and the third by Wilkins and by Franklin. So Franklin got to say, in Nature, in the same issue as Watson and Crick, everything she knew about DNA, including the publication of the photographs."

Elkin, however, points out the following:

> "In the early years, only crystallographers knew that there was something wrong with the three papers. They looked at their brilliant, brilliant first [Watson and Crick] paper. And I want to be clear, I think everybody deserves credit for this. I do not diminish anybody's contribution to this work. And then it's, okay, where's the data? Anybody who knew anything knew we need data for this.
>
> Then we look at Wilkin's paper and say they're as near. Then they look at Franklin's paper and there is the data but why isn't she acknowledged? And part of that was because of the snaky deal that was done between Randall, the head of Kings's [where Rosalind Franklin and Wikens worked independently], and Bragg, the head of Cavendish [where Watson and Crick worked], to cover up the very awkward fact that the data had migrated from one place to the other."

In the 1953 Watson and Crick manuscript itself, the authors have the following to say:

> "We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London."
>
> and
>
> "So far as we can tell, [our model for DNA structure] is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications [from Dr. Wilkins and Dr. Franklin]. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments."

Two more notes about these statements are made in the previously noted commentary, however:

> 'Watson and Crick say that they "were not aware of the details" of the work of King's College scientist Rosalind Franklin – a statement that marks what many consider an inexcusable failure to give Franklin proper credit. According to Lynne Elkin ... Watson and Crick were not aware of all the details of Franklin's work, but they were aware of enough of the details to discover the structure of DNA. Yet this paper does not ever formally acknowledge her, instead concealing

her significant role by saying they "were not aware" of her work ... Franklin was at work on the DNA puzzle using X-ray crystallography, which involved taking X-ray photographs of DNA samples to infer their structure. By late February 1953, her analysis of these photos brought her close to the correct DNA model.'

and

'Interestingly... a stronger acknowledgment of Franklin's work [appears] in an early draft of the paper: "We have also been stimulated by the very beautiful experimental work of Dr. M. H. Wilkins and his co-workers at Kings College, London." Elkin suggests that the phrase "very beautiful" is most likely a nod to Franklin's X-ray photograph. The same draft also acknowledged Franklin's work with the sentence: "It is known that there is much unpublished experimental material." When Maurice Wilkins read the draft, he advised Watson and Crick to delete this sentence and the phrase "very beautiful." They agreed to his suggestion.'

In your responses to the questions below, consider using hedging to try to make your arguments more approachable in a manner that invites consideration, engagement, and exploration. This will likely have a better chance of "winning someone over" as opposed to merely expressing an opinion *since (as you probably know) what they say about opinions is... everybody has one.*

(a) What are your impression so far? Are you amenable to the Watson and Crick position that they used publicly available information to create the double-helix model and appropriately acknowledged their sources? Or, is their representation of their reliance on Rosalind Franklin's data not sufficient and in actual fact simply plagiarism?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Since Wilkins, Crick, and Watson (but not Rosalind Franklin) were award the Nobel Prize, Rosalind Franklin has become universally accepted as deserving credit in the discovery of the double-helix structure of DNA. Do you think she has finally recieved proper acknoweledgement and credit? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 8: Conflicts of Interest

While understanding the relevance of Rosalind Franklin's data to their thesis and having access to this data, Watson and Crick produced a manuscript without Rosalind Franklin as a co-author. Indeed, Watson and Crick never even communicated with Rosalind Franklin about her data directly at any point. Still, in STA130 professor Scott Schwartz's assessment, since Watson and Crick did not present (and only referenced) Franklin's **data** in their manuscript (which is universally accepted as having merit and was published on that basis), and only suggested its alignment with their proposed theory, the acknowledgments they provided in their paper were "acceptable and sufficient". This is a much more optimistic assessment than that of preeminent Rosalind Franklin scholar Lynne Osman Elkin, who argues that the Watson and Crick manuscript intentionally obfuscates the truth of Rosalind Franklin's contribution. Again from the previously noted commentary:

'Then in perhaps the most pivotal moment in the search for DNA's structure, Wilkins, a longtime friend of Crick, showed Watson one of Franklin's photographs without Franklin's permission. Watson recalled, "The instant I saw the picture my mouth fell open and my pulse began to race." To Watson, the cross-shaped pattern of spots in the photo meant that DNA had to have a helical structure. Franklin's photograph was critical in solving the problem, as Watson admitted in his 1968 book, The Double Helix. Watson and Crick also had access to an internal report from the Medical Research Council, a British agency for funding life sciences, summarizing much of Franklin's unpublished work on DNA, including precise measurements of the molecule. As the Cavendish representative to the agency, scientist Max Perutz had a copy of the report, and when Crick asked to see it, Perutz obliged. While the report was not confidential, science historian

> Lynne Elkin contends that "showing unpublished work to an unacknowledged competitor was a questionable act which justifiably infuriated" John Randall, the head of King's. Crick later said the data in the report enabled him to reach the significant conclusion that DNA has two chains running in opposite directions. Although Franklin was listed in the acknowledgements section with other scientists, there was no specific mention of her contributions.'

Wilkins and Rosalind Franklin worked independently of each other at King's College. And Watson and Crick worked at the "competitor" Cavendish laboratory. So why did Wilkins have Rosalind Franklin's image and decide to share with Crick without Rosalind Franklin's permission? From the commentary and Brenda Maddox's Nature article:

> "Wilkins, [was] a longtime friend of Crick ... Franklin decided to take another job. As she was preparing to leave, she turned her X-ray photographs over to her colleague Maurice Wilkins."
>
> "[Rosalind Franklin] was, in fact, so unhappy at King's that, in early 1953, getting out as fast as possible was far more important to her than finishing her work on DNA."
>
> This meant that Rosalind Franklin's project reverted to Wilkins, who's own work in this area was ongoing, and who did not wish to maintain connections with Rosalind Franklin and subsequently seemingly interpreted himself to have agency and determination over her materials.

What about Max Perutz? From wikipedia:

> Rosalind Franklin's images and data were submitted as part of a "Medical Research Council report" under the management of John Randall, and were subsequently shared with Cavendish representative Max Perutz. Perutz subsequently gave Rosalind Franklin's images to Crick without consulting Rosalind Franklin (or John Randall). Perutz later stated that (a) the images were part of a King's College portfolio which he was told should be shared with other research groups in a "non confidential manner" and that (b) anyway, Rosalind Franklin gave a public talk with the same images which Watson had attended so they were already "public knowledge".

In your responses to the questions below, be mindful of opportunities to use hedging to try to make your arguments more approachable in a manner that invites consideration, engagement, and exploration.

(a) In the context of the "race to the double helix", in which the Cavendish and King's groups were viewed as competitors, what do you make of the choice by Watson and Crick to benefit off of the work of Rosalind Franklin without approaching directly (let alone engaging her as a collaborator)?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Should we accept the acknowlegements of the Watson and Crick manuscript (noted in Question 7 above) as "acceptable and sufficient"? Should we be hesitant to take them at face value and trust that they are indeed "acceptable and sufficient"?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) What do you make of the fact that Wilkins shared Rosalind Franklin's photographs with Crick without consulting Rosalind Franklin? Explain why you find, or do not find Wilkin's actions reasonable or understandable.

REPLACE THIS TEXT WITH YOUR ANSWER

(d) What do you make of the fact that Perutz shared a report "summarizing much of Franklin's unpublished work on DNA, including precise measurements [and images] of the molecule" with Crick without consulting Rosalind Franklin? Explain why you find, or do not find Perutz's actions reasonable or understandable.

REPLACE THIS TEXT WITH YOUR ANSWER

(e) Wilkins, Crick, Watson AND Franklin are now understood to deserve credit for the discovery of the double-helix structure of DNA. Watson and Crick had Wilkins review their draft manuscript, and made edits to their final manuscript as a result of Wilkins recommendations. What do you make of the fact that Watson and Crick did not ask Rosalind Franklin to review and provided feedback as they finalized their manuscript? Explain why you find, or do not find their actions reasonable or understandable.

REPLACE THIS TEXT WITH YOUR ANSWER

Question 9: Personal Decisions

John Randall, the head of King's college where Rosalind Franklin worked was reportedly furious at Perutz for "showing unpublished work to an unacknowledged competitor". Nonetheless, "The authors of the three papers [Wilkins, Crick, Watson, and Franklin], their lab chiefs, and the editors of Nature agreed that all three would be published in the same issue." New York Times science reporter Nicholas Wade and preeminent Rosalind Franklin scholar Lynne Osman Elkin debated this point in their panel discussion:

> Wade: "[T]he problem with that saying Rosalind was ill-treated is that there's absolutely no evidence that she herself believed this to be the case."
>
> Elkin: "She didn't know."
>
> Wade: "She was definitely in a position to complain if she wished. She had just arranged a new job. She was leaving the King's College department to go to Birkbeck College. We know that she complained vociferously about things she thought were unfair, like being paid less than – at the MRC – at being paid less than men who did the same job. But she never, ever, complained about this. Moreover, she became great close friends with Watson and with Crick. ... unlikely – if in fact she felt they had stolen her discovery. She must have known that they were using her data because there were no other data – her data are acknowledged in Crick's paper."

(a) Given the often encountered representation that the Nobel Prize was "stolen" from Rosalind Franklin, and the now universal consensus of the relevance and contribution of Rosalind Franklin's work, do you think it's possible that Rosalind Franklin could (or should?) have been more assertive in seeking recognition for her work? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Assuming Rosalind Franklin should have sought more recognition for her work during her lifetime, what reasons do you think might have contributed to her not doing so?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) What circumstances do you think lead to neither Watson and Crick, Wilkins and nor John Randall nor Max Perutz, nor the relevant lab chiefs and the editors of Nature to stand up for a greater acknowledgement of Rosalind Franklin at the critical points leading up to the 1953 publication?

REPLACE THIS TEXT WITH YOUR ANSWER

(d) What could be happening that so many people never communicated with Rosalind Franklin directly to discuss the relevance and implications of her data?

REPLACE THIS TEXT WITH YOUR ANSWER

(e) None Watson and Crick's use of Rosalind Franklin's data involved any explicit consultation with Rosalind Franklin, and it could have. How much of Watson and Cricks decisions here do you think boiled down to "It's fine no need to check", "We didn't even think about it", and/or "We should maybe not do this. . . "?

REPLACE THIS TEXT WITH YOUR ANSWER

(f) How much of Maurice Wilkins and Max Puretz decisions to share their access to Rosalind Franklin's data boiled down to "It's fine no need to check", "We didn't even think about it", and/or "We should maybe not do this. . . "?

REPLACE THIS TEXT WITH YOUR ANSWER

- *Note: Elkin and Wade in their panel discussion appear to agree that Crick (in subsequent manuscripts and statements) and Watson (in his memoir and subsequent statements) did in fact work to ensure Rosalind Franklin was acknowledged and recognized. It is generally accepted that the comments from Watson and Crick were crucial for raising awareness of the contribution and relevance of Rosalind Franklin's work. . . . but they took many, many years to do so. . . and did not do so during Rosalind Franklin's lifetime.*

Question 10: Sexism

We've seen that Rosalind Franklin was moving jobs. From the commentary:

> "What exactly was Franklin's research, and how did Watson and Crick gain access to it? While they were busy building their models, Franklin was at work on the DNA puzzle using X-ray crystallography, which involved taking X-ray photographs of DNA samples to infer their structure. By late February 1953, her analysis of these photos brought her close to the correct DNA model.
>
> But Franklin stopped her work on DNA because she was frustrated with a strained environment at King's, one that pitted her against her colleagues. In an institutional culture that barred women from the dining room and other social venues, she was denied access to the informal discourse that is essential to any scientist's work. Seeing no chance for a tolerable professional life at King's, Franklin decided to take another job. As she was preparing to leave, she turned her X-ray photographs over to her colleague Maurice Wilkins."

Indeed, one need only give a cursory glance of the account of Rosalind Franklin given by Watson in The Double Helix to understand the prevalence of gender bias and stereotypes (and outright sexism and misogyny) which permeated King's.

However, Brenda Maddox finds that there's more to story as well:

> "Women's exclusion from the King's senior common room deprived Franklin of the intellectual companionship of her colleagues[?] In fact, most of the scientific staff preferred to eat in the joint dining room, men and women together, and the women, in general, felt well treated at King's. . . . I found [Rosalind Franklin to be] a more attractive, capable woman than Watson had suggested, and a King's College more congenial and welcoming to women scientists than Sayre had allowed. I also found that Franklin felt singularly unhappy at King's, not so much because of her gender, but because of her class and religion: a wealthy Anglo-Jew felt out of place in a Church of England setting dominated by swirling cassocks and students studying for the priesthood."At King's," she wrote to Sayre (albeit inaccurately), "there are neither Jews nor foreigners".

From a nature profile:

> 'Friends and close colleagues considered Franklin a brilliant scientist and a kindhearted woman. However, she could also be short-tempered and stubborn, and some fellow scientists found working with her to be a challenge. Among them was Maurice Wilkins, the man she was to work with at King's College.

> A misunderstanding resulted in immediate friction between Wilkins and Franklin, and their clashing personalities served to deepen the divide. The two were to work together on finding the structure of DNA, but their conflicts led to them working in relative isolation. While this suited Franklin, Wilkins went looking for company at "the Cavendish" laboratory in Cambridge where his friend Francis Crick was working with James Watson on building a model of the DNA molecule.'

And from another article:

> "Wilkins was out of town when Rosalind Franklin was hired, and their first meeting did not go well. He mistook her for a new secretary, and it appears that they never established a good working relationship as fellow scientists. Franklin had gotten the impression she would be working alone in the lab; Wilkins saw her in a more supportive role to his work. In Wilkins' later memoir, he referred to Franklin as 'Rosy', complaining that she had an attitude and was too independent."

The name "Rosy" was not used by Rosalind Franklin, but was the name Wilkins and Watson apparently preferred for talking about her, and is generally taken to be misogynistic (if not derogatory and pejorative). The same article above goes on to report that 'Watson "did not like Franklin" ', but this is a gross oversimplification of the characterization Watson gives to Rosalind Franklin and his later continued relationship with Rosalind Franklin in his memoir. Nonetheless, all of this sheds light on why Rosalind Franklin did not share the same collegial relationship with Watson and Crick and Wilkins that they themselves enjoyed.

(a) Does the kind of environment Rosalind Franklin was working seem like one in which she might not be given appropriate recognition and acknowledgement? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) How much of Rosalind Franklin's experience do you feel boils down to "personality" versus "sexism"?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) How much of the fact that Rosalind Franklin was not consulted and kept "in the loop" do you feel can be explained by the timing of her change of position and research project as opposed to other explanations?

REPLACE THIS TEXT WITH YOUR ANSWER

(d) What possible changes to the contextual situation in which Rosalind Franklin found herself would have been conducive to her receiving a more appropriate level of recognition and acknowlegement for her work?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 11: Modern "Big Team Science" Authorship

STA130 Head Teaching Assistant (managing 24 Teaching Assistants) Quin Xie comments:

> "Whether or not it's recognition or authorship under debate, Rosalind Franklin's contribution was not sufficiently recognized in academia or publically until many years after her death."
>
> "The standard for attributing authorship and the awareness of discrimination may have changed over the years. It is common for consortium-based studies to include people as authors just for collecting and providing patient data. Any PI (Principle Investigator) whose intellectual input influenced the progress of the project would be included as an author nowadays, and it is known that Watson and Crick could not solve the structure of DNA until they saw (without explicit consent) the preliminary data from Rosalind Franklin."

(a) Do you think Rosalind Frankin's "public presentation" and the distribution of her images and data through the inter-laboratory "MRC update report" constitute a "consortium-based study" or "collecting and providing study data"? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Assuming there will be proper acknowledgement, do you think presenting ones "unpublished" work in public or submitting your material in a report intended for distribution gives sufficient explicit consent for its use (even if a mistake)? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) Assuming there will be proper acknowledgement, do you think presenting ones "unpublished" work in public or submitting your material in a report intended for distribution is gives sufficient "implicit" consent for its use (even if not explicit consent)? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(d) If one attends to a talk or reads a distributed report and then makes a reference to the data and ideas therein, do you find it problematic that if the material is "used" without asking the (primary) originator for explicit permission to do so and confirming with them the adequacy of your acknowledgement provided in the "use"? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(e) Some of Rosalind Franklin's work and data were "publically shared" and so available for Watson and Crick to use so long as they acknowledge and reference it appropriately; however, each additional iota of data and information sought and acquired by Watson and Crick beyond this is intellectual fraud, plagiarism, and stealing of intellectual property. Do you agree with all or some of this statement? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(f) Having arrived at the end of Part 3, consider the following questions:

- Do you believe Rosalind Franklin was specifically not extended authorship on gender grounds in some degree?
- Do you believe Maurice Wilkins, who like Rosalind Franklin was not extended authorship, but who was awarded the Nobel prize with Watson and Crick in 1962, was less deserving of being extended authorship than Rosalind Franklin?
  - Note: In securing Wilkins Nobel Prize bid, Crick stated: "He deserves it. He did very, very important work getting the structure initially, doing preliminary work, and at the end doing brilliant work confirming it. But the actual data he used was that of Rosalind Franklin. Period."
- Do you believe Raymond Gossling who was Rosalind Franklin's PhD student and who was sufficiently experimentally instrumental to co-author the work (and data and images) in question, but was not extended authorship was less deserving of being extended authorship and given recognition than Rosalind Franklin?
- Do you find that the co-authors of Maurice Wilkins, both of whom were not extended authorship, were less deserving of being extended authorship than Rosalind Franklin?

Based on everything you've read, and the authorship assignment patters implicit in the above questions, summarize your thoughts about whether or not "...[Rosalind Franklin] was wrongfully denied authorship on the original Crick and Watson paper" and why you think "Franklin's pivotal contribution to the discovery of the structure of DNA initially went unrecognized..."

REPLACE THIS TEXT WITH YOUR ANSWER