Homework 2

Ryo Iwata

26 February, 2024

Homework 2

Instructions:

0. Start early and leave time to review and revise your submission well in advance of the due date.
1. Read the problem in its entirety before starting. Often, we can ignore certain variables while working on early problems and then pay attention to them in later problems. Students can waste time trying to incorporate these effects too early and with limited guidance on how to do this.
2. Plan your approach to tidying the data and preparing it for figures and analysis.
3. Translate your plan into well structured comments (in your code chunks)
4. Translate your comments into code, leaving the comments in place (so we understand your intent)
5. Ensure your code, figure and text answers are neat and tidy, grammatically correct, clear, and concise. You may find the reindent lines (ctrl-i) or reformat code (ctrl-A i.e. ctrl-shift-a) useful. Break arguments to functions onto different lines and ensure your code and comments do not flow off the edge of the page.
6. Review your output (pdf) to ensure your outputs render well.
7. All tables, formulae, etc., should be carefully typeset. Do not just dump outputs from functions.

Note: We will grade both your answer and how you arrived at said answer (i.e., your code). Be sure to use sensible and sufficiently descriptive tibble, variable and model names. Make efficient and appropriate use of the tidyverse functionality.

Problem #1 (60 pts):

Your laboratory is evaluating the diffusion of drug delivery vehicles (particles) injected into the brain. The lab is seeking to understand how the diameter *and* surface charge of the drug delivery vehicle affects the vehicle's diffusion.

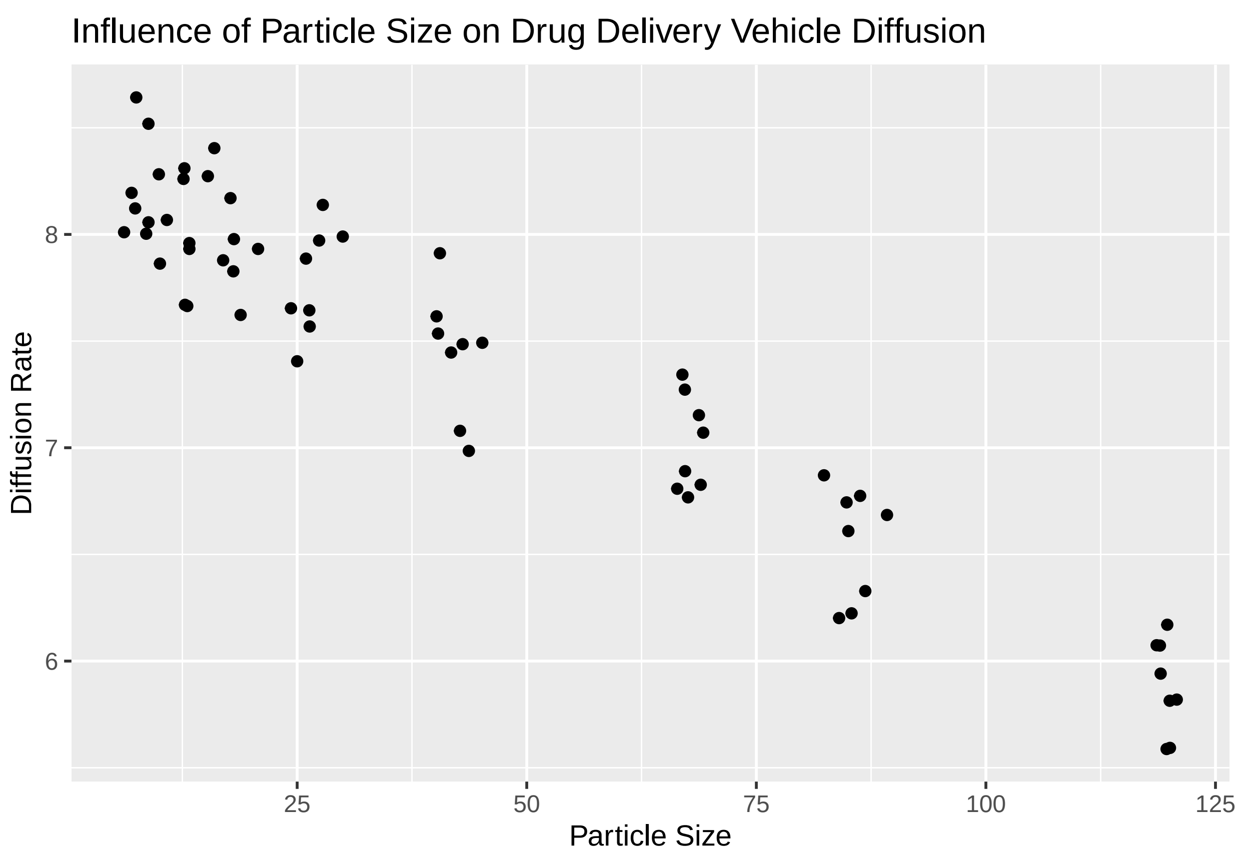
Load and evaluate the structure of data in HW2Data sheet DiffSizeCharge (only workbook/sheet in file)

```
drug_diffusion_raw <- readxl::read_excel("./data/HW2data.xlsx", sheet="DiffSizeCharge")
```

1) Investigate Particle Size and Diffusion Rate. At this point, ignore the effects of surface charge.

a) Create a plot to visualize the effect of particle diameter on particle diffusion. Think about the regression model you may be fitting, and place the appropriate variables on the x axis and y axis. Do not include any summary or smooth geometries (2 pts)

```
# Plotting scatter plot of Particle size against Diffusion rate
drug_diffusion_raw |> ggplot(aes(x=`Particle Size`, y=`Diffusion Rate`)) +
  geom_point() +
  labs(title = "Influence of Particle Size on Drug Delivery Vehicle Diffusion")
```



b) Fit a univariate (1-variable) regression model to these data. (4 pts)

```
# Regress diffusion rate onto particle size
size_to_diffusion_model <- lm(`Diffusion Rate` ~ `Particle Size`, drug_diffusion_raw)
```

c) What are the model's estimated coefficients? Write out the generic formula for this model (i.e., with β terms) and then again with the estimates correctly substituted (4 pts)

```r
tidy_model_size_to_diffusion <- tidy(size_to_diffusion_model)

kable(tidy_model_size_to_diffusion)
```

term	estimate	std.error	statistic	p.value
(Intercept)	8.3279	0.0489	170.23	0
`Particle Size`	-0.0204	0.0008	-25.18	0

```r
intercept_size_to_diffusion_model <- pull(filter(tidy_model_size_to_diffusion,
                                                  term=="(Intercept)"), estimate)

slope_size_to_diffusion_model <- pull(filter(tidy_model_size_to_diffusion,
                                              term=="`Particle Size`"), estimate)

print(paste("Coefficient for intercept: ", intercept_size_to_diffusion_model))
```

```
## [1] "Coefficient for intercept:  8.32792802062674"
```

```r
print(paste("Coefficient for Particle Size: ", slope_size_to_diffusion_model))
```

```
## [1] "Coefficient for Particle Size:  -0.0203730765105808"
```

Answer: Generic formula for this model $\hat{y} = \beta_0 + \beta_{diameter} x_{diameter}$

Answer: Formula for this model with the estimates substituted $\hat{y} = 8.3279 + -0.0204 x_{diameter}$

d) Calculate and present in a table the $\mathbf{SS_x}$, $\mathbf{SS_y}$, $\mathbf{SS_{xy}}$, $\mathbf{SS_{Regression}}$, and $\mathbf{SS_{Error}}$ for this model. (10 pts)

```r
# Adding lm information to data
augment_size_to_diffusion_model <- augment(size_to_diffusion_model)

# Testing if the regression coefficients are equal to zero
tidy_anova_size_to_diffusion <- tidy(car::Anova(size_to_diffusion_model))
```

```r
# Calculating the sum of squares of X, the predictor variable
SSx_size_to_diffusion <- pull(summarise(augment_size_to_diffusion_model,
                          SSx = sum((`Particle Size` - mean(`Particle Size`))^2)
                          )
            )

# Calculating the sum of squares of error/residual
# which measures the discrepancy between the data and the model)
SSerror_size_to_diffusion <- pull(filter(tidy_anova_size_to_diffusion,
                                          term=="Residuals"), sumsq)

# Calculating the sum of squares of regression
# or how well the model represents the data
SSregression_size_to_diffusion <- pull(filter(tidy_anova_size_to_diffusion,
                                              term=="`Particle Size`"), sumsq)

# Calculating the sum of squares of Y, the predictor variable
```

```r
SSy_diffusion <- SSerror_size_to_diffusion + SSregression_size_to_diffusion

# Calculating the sum of squares of X and Y
SSxy_size_to_diffusion <- pull(summarise(augment_size_to_diffusion_model,
                         SSxy = sum((`Particle Size` - mean(`Particle Size`)) *
                                      (`Diffusion Rate` - mean(`Diffusion Rate`))
                                    )
                         )
                         )

kable(tibble(term = c("SSx", "SSy", "SSxy", "SSRegression", "SSerror"),
             value = c(SSx_size_to_diffusion, SSy_diffusion,
                       SSxy_size_to_diffusion, SSregression_size_to_diffusion,
                       SSerror_size_to_diffusion)
             )
      )
```

term	value
SSx	89746.034
SSy	40.893
SSxy	-1828.403
SSRegression	37.250
SSerror	3.643

e) What is R^2 for this model (2 pts)?

```r
# Calculating R^2, which is another measure of fit that is between 0 and 1
# Which is the proportion of the variability of Y that can be explained by X
r_squared_size_to_diffusion <- (SSy_diffusion - SSerror_size_to_diffusion) / SSy_diffusion

print(paste("R-squared for the model: ", r_squared_size_to_diffusion))
```

```
## [1] "R-squared for the model:  0.91091452224952"
```

f) Using the estimated slope term (β_1) and its associated error, use a t-test to evaluate whether the slope term is zero or non-zero. What is the t-statistic and associated p-value (2 pts)?

```r
size_to_diffusion_tstat <- pull(filter(tidy_model_size_to_diffusion, term=="`Particle Size`"),
                         statistic)

size_to_diffusion_t_pvalue <- pull(filter(tidy_model_size_to_diffusion, term=="`Particle Size`"),
                          p.value)

print("t-statistic and associated p-value for slope term")
```

```
## [1] "t-statistic and associated p-value for slope term"
```

```r
print(paste("T(", df.residual(size_to_diffusion_model), "):",
            size_to_diffusion_tstat,
            "p = ", size_to_diffusion_t_pvalue))
```

```
## [1] "T( 62 ): -25.1785781155024 p =  2.9349951753222e-34"
```

Answer: Assuming significance with a p-value less than 0.05, we can reject the null hypothesis that the slope term is zero. Which is evidence suggesting a significant linear relationship between the independent and dependent variables in the model

g) Does particle size have a significant effect on particle diffusion with this model? (2 pts)

```r
tidy_model_size_to_diffusion
```

```
## # A tibble: 2 x 5
##   term            estimate std.error statistic  p.value
##   <chr>              <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)       8.33    0.0489       170.  1.64e-84
## 2 `Particle Size`  -0.0204  0.000809     -25.2 2.93e-34
```

```r
kable(tidy_anova_size_to_diffusion)
```

term	sumsq	df	statistic	p.value
`Particle Size`	37.250	1	634	0
Residuals	3.643	62	NA	NA

```r
size_to_diffusion_fstat <- pull(filter(tidy_anova_size_to_diffusion, term=="`Particle Size`"),
                                statistic)

size_to_diffusion_pvalue <- pull(filter(tidy_anova_size_to_diffusion, term=="`Particle Size`"),
                                 p.value)

dfnum_size_to_diffusion <-

print("F-statistic and associated p-value for all slope terms")
```

```
## [1] "F-statistic and associated p-value for all slope terms"
```

```r
print(paste("F(1,", df.residual(size_to_diffusion_model), ") = ",
            size_to_diffusion_fstat, ", p = ",
            size_to_diffusion_pvalue))
```

```
## [1] "F(1, 62 ) =  633.960795918457 , p =  2.93499517532216e-34"
```

Answer: Assuming significance with a p-value less than 0.05, we can reject the null hypothesis that all slope coefficients in the model are equal to zero because the p-value for the F-statistic is lower than the threshold. This is evidence that the model fits the data significantly better than a model without predictors.

```r
kable(tidy_model_size_to_diffusion)
```

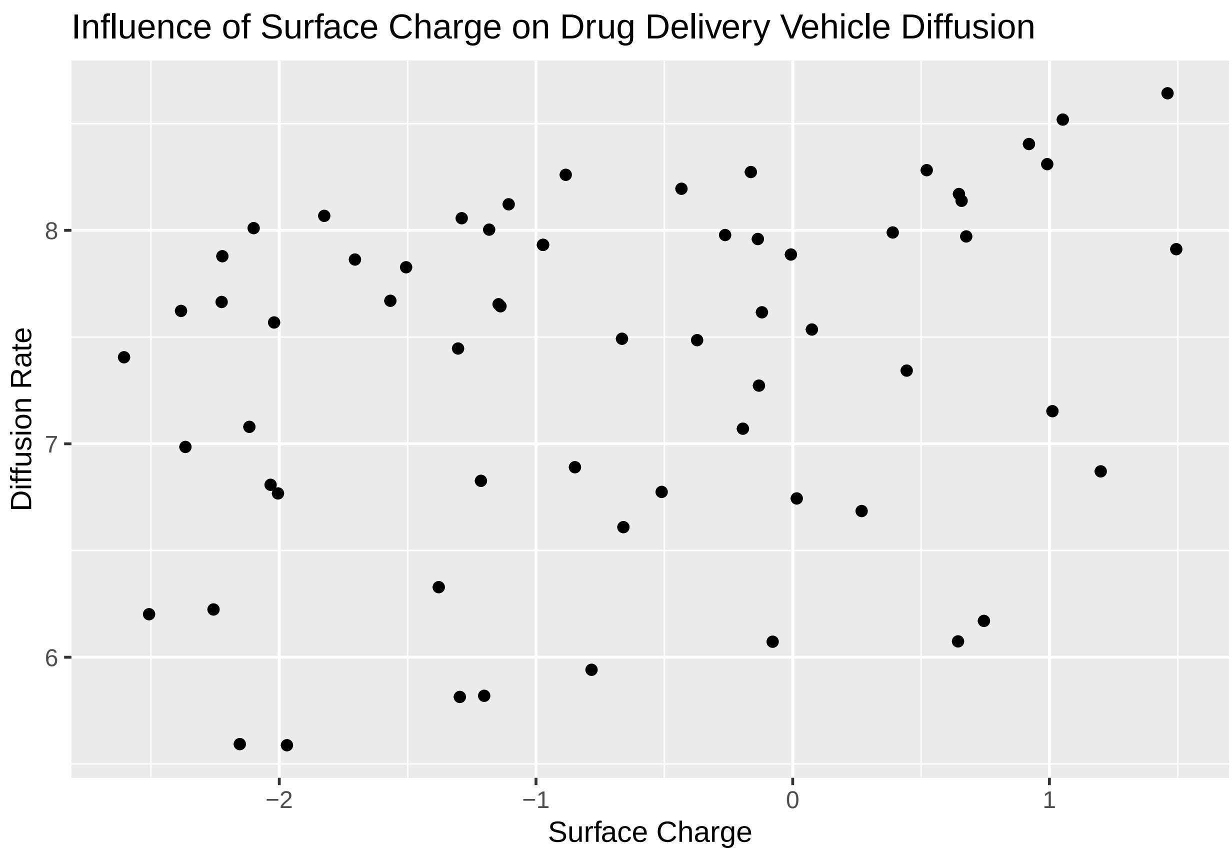
term	estimate	std.error	statistic	p.value
(Intercept)	8.3279	0.0489	170.23	0
`Particle Size`	-0.0204	0.0008	-25.18	0

We can also reject the null hypothesis that the coefficient associated with particle size is equal to zero because the p-value for the t-statistic is lower than the threshold as shown in question `1.f`. The analysis suggests that particle size has a significant effect on particle diffusion because we have evidence that there is a significant linear relationship between particle size and particle diffusion.

2) In the same experiment, these particles had different surface charges. You decide to investigate the effect of surface charge on particle diffusion.

a) Create a plot to visualize the effect of surface charge on particle diffusion in whole blood. At this point, ignore the effect of particle diameter (2 pts).

```
drug_diffusion_raw |> ggplot(aes(x=`Surface Charge`, y=`Diffusion Rate`)) +
  geom_point() +
  labs(title = "Influence of Surface Charge on Drug Delivery Vehicle Diffusion")
```



b) Fit a univariate (1-variable) regression model to these data. (4 pts)

```
# Regress diffusion rate onto surface charge
charge_to_diffusion_model <- lm(`Diffusion Rate` ~ `Surface Charge`, drug_diffusion_raw)
```

c) What are the model's estimated coefficients? Write out the formula for the model with generic terms as well as substituting the estimated coefficients (2 pts)

```
tidy_model_charge_to_diffusion <- tidy(charge_to_diffusion_model)

kable(tidy_model_charge_to_diffusion)
```

term	estimate	std.error	statistic	p.value
(Intercept)	7.5015	0.1144	65.564	0.0000
Surface Charge	0.2009	0.0859	2.338	0.0226

```r
intercept_charge_to_diffusion_model <- pull(filter(tidy_model_charge_to_diffusion,
                                                   term=="(Intercept)"), estimate)
slope_charge_to_diffusion_model <- pull(filter(tidy_model_charge_to_diffusion,
                                               term=="`Surface Charge`"), estimate)

print(paste("Coefficient for intercept: ", intercept_charge_to_diffusion_model))
```

```
## [1] "Coefficient for intercept:  7.50148869502308"
```

```r
print(paste("Coefficient for Surface charge: ", slope_charge_to_diffusion_model))
```

```
## [1] "Coefficient for Surface charge:  0.200906975903284"
```

Answer: Generic formula for this model $\hat{y} = \beta_0 + \beta_{charge} x_{charge}$

Answer: Formula for this model with the estimates substituted $\hat{y} = 7.50148 + -0.20090 x_{charge}$

d) Calculate and present in a table the $\mathbf{SS_x}$, $\mathbf{SS_y}$, $\mathbf{SS_{xy}}$, $\mathbf{SS_{Regression}}$, and $\mathbf{SS_{Error}}$ for this model. (10 pts)

```r
# Adding lm information to data
augment_charge_to_diffusion_model <- augment(charge_to_diffusion_model)

# Testing if the regression coefficients are equal to zero
tidy_anova_charge_to_diffusion <- tidy(car::Anova(charge_to_diffusion_model))
```

```r
# Calculating the sum of squares of X, the predictor variable
SSx_charge_to_diffusion <- pull(summarise(augment_charge_to_diffusion_model,
                      SSx = sum((`Surface Charge` - mean(`Surface Charge`))^2)
                      )
            )

# Calculating the sum of squares of error/residual
# which measures the discrepancy between the data and the model)
SSerror_charge_to_diffusion <- pull(filter(tidy_anova_charge_to_diffusion,
                                           term=="Residuals"), sumsq)

# Calculating the sum of squares of regression
# or how well the model represents the data
SSregression_charge_to_diffusion <- pull(filter(tidy_anova_charge_to_diffusion,
                                           term=="`Surface Charge`"), sumsq)

# Calculating the sum of squares of X and Y
SSxy_charge_to_diffusion <- pull(summarise(augment_charge_to_diffusion_model,
                       SSxy = sum((`Surface Charge` - mean(`Surface Charge`)) *
                                    (`Diffusion Rate` - mean(`Diffusion Rate`))
                                  )
                       )
                       )

kable(tibble(term = c("SSx", "SSy", "SSxy", "SSRegression", "SSerror"),
             value = c(SSx_charge_to_diffusion, SSy_diffusion,
                       SSxy_charge_to_diffusion, SSregression_charge_to_diffusion,
                       SSerror_charge_to_diffusion)
             )
      )
```

term	value
SSx	82.063
SSy	40.893
SSxy	16.487
SSRegression	3.312
SSerror	37.581

e) What is R^2 for this model (2 pts)?

```
# Calculating R^2, which is another measure of fit that is between 0 and 1
# Which is the proportion of the variability of Y that can be explained by X
r_squared_charge_to_diffusion <- (SSy_diffusion - SSerror_charge_to_diffusion) / SSy_diffusion

print(paste("R-squared for the model: ", r_squared_charge_to_diffusion))
```

```
## [1] "R-squared for the model:  0.0810007112747673"
```

f) Using the estimated slope term (β_1) and its associated error, use a t-test to evaluate whether the slope term is zero or non-zero. What is the t-statistic and associated p-value (2 pts)?

```
charge_to_diffusion_tstat <- pull(filter(tidy_model_charge_to_diffusion, term=="`Surface Charge`"),
                       statistic)

charge_to_diffusion_t_pvalue <- pull(filter(tidy_model_charge_to_diffusion, term=="`Surface Charge`"),
                        p.value)

print("t-statistic and associated p-value for slope term")
```

```
## [1] "t-statistic and associated p-value for slope term"
```

```
print(paste("T(", df.residual(charge_to_diffusion_model), "):",
            charge_to_diffusion_tstat,
            "p = ", charge_to_diffusion_t_pvalue))
```

```
## [1] "T( 62 ): 2.33766714669049 p =  0.0226471384851704"
```

Answer: Assuming significance with a p-value less than 0.05, we can reject the null hypothesis that the slope term is zero. Which is evidence suggesting a significant linear relationship between the independent and dependent variables in the model

g) Does particle size have a significant effect on particle diffusion in this model? (2 pts)

```
kable(tidy_anova_charge_to_diffusion)
```

term	sumsq	df	statistic	p.value
`Surface Charge`	3.312	1	5.465	0.0226
Residuals	37.581	62	NA	NA

```
charge_to_diffusion_fstat <- pull(filter(tidy_anova_charge_to_diffusion, term=="`Surface Charge`"),
                       statistic)
```

```r
charge_to_diffusion_pvalue <- pull(filter(tidy_anova_charge_to_diffusion, term=="`Surface Charge`"),
                        p.value)

dfnum_charge_to_diffusion <-

print("F-statistic and associated p-value for all slope terms")
```

```
## [1] "F-statistic and associated p-value for all slope terms"
```

```r
print(paste("F(1,", df.residual(charge_to_diffusion_model),
            ") = ", charge_to_diffusion_fstat,
            ", p = ", charge_to_diffusion_pvalue))
```

```
## [1] "F(1, 62 ) =  5.46468768871605 , p =  0.0226471384851704"
```

Answer: Assuming significance with a p-value less than 0.05, we can reject the null hypothesis that all slope coefficients in the model are equal to zero because the p-value for the F-statistic is lower than the threshold. This is evidence that the model fits the data significantly better than a model without predictors.

```r
kable(tidy_model_charge_to_diffusion)
```

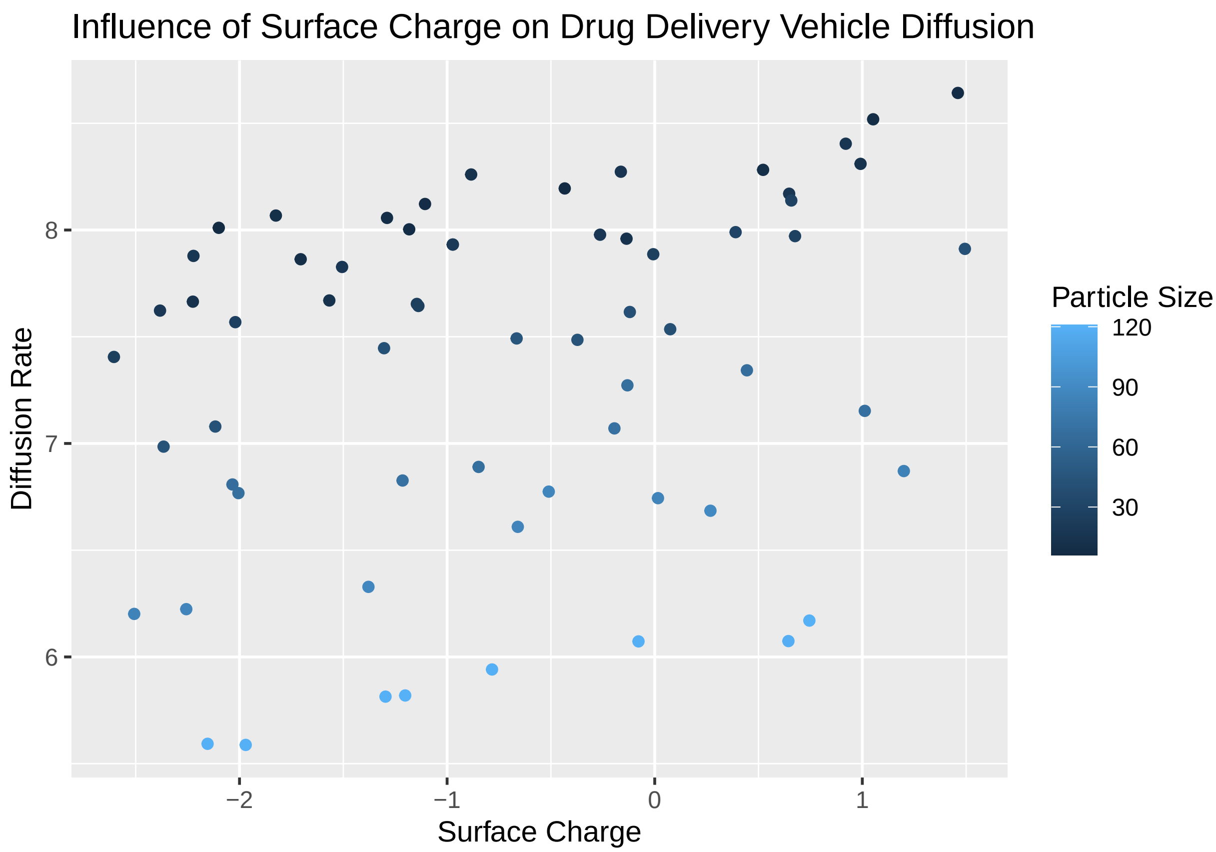
term	estimate	std.error	statistic	p.value
(Intercept)	7.5015	0.1144	65.564	0.0000
`Surface Charge`	0.2009	0.0859	2.338	0.0226

We can also reject the null hypothesis that the coefficient associated with surface charge is equal to zero because the p-value for the t-statistic is lower than the threshold as shown in question `2.f`. The analysis suggests that surface charge has a significant effect on particle diffusion because we have evidence that there is a significant linear relationship between surface charge and particle diffusion.

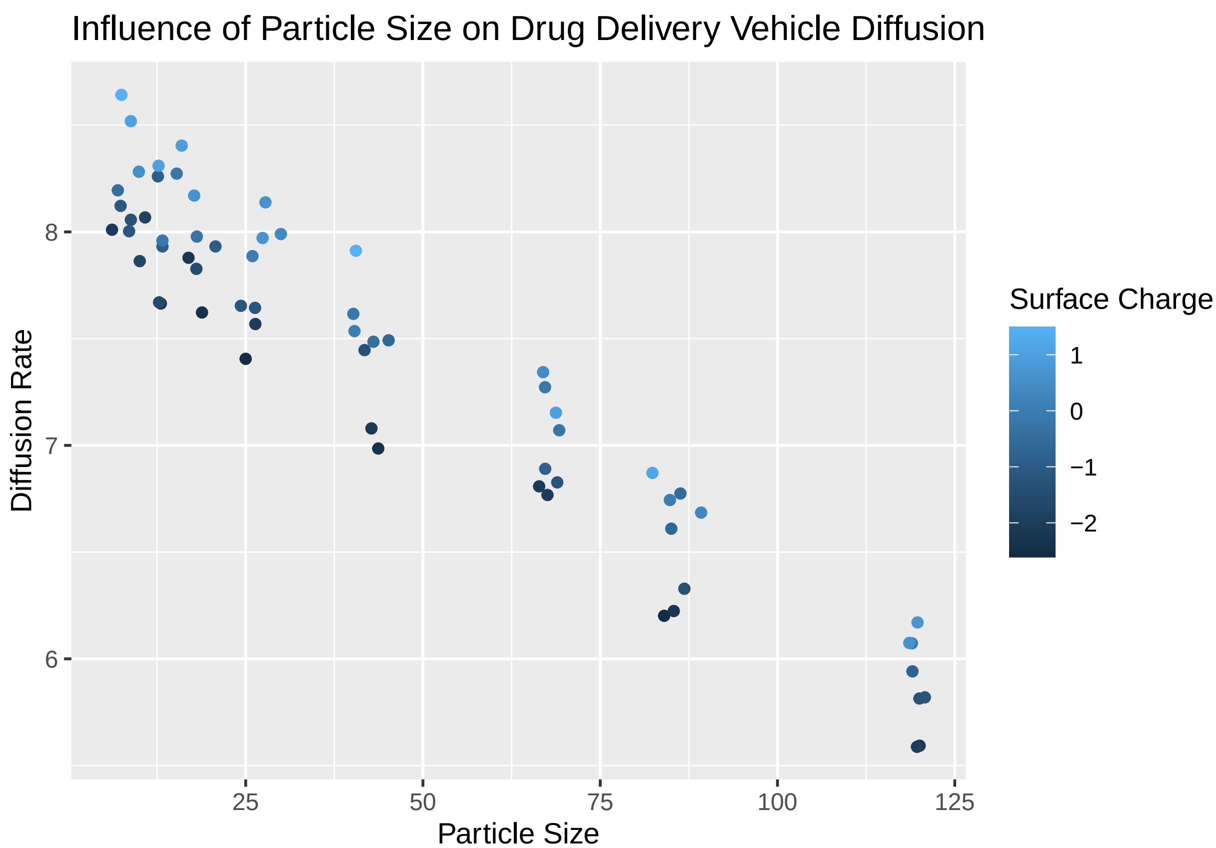
3) Multiple Regression

a) Create a 2-D scatterplot to visualize combined effect of surface charge and particle diameter on particle diffusion in whole blood. Think about the regression model you may be fitting, and use appropriate aesthetic mappings. Do not include summary or smooth geoms/stats (3 pts).

```r
drug_diffusion_raw |> ggplot(aes(x=`Surface Charge`, y=`Diffusion Rate`, color=`Particle Size`)) +
  geom_point() +
  labs(title = "Influence of Surface Charge on Drug Delivery Vehicle Diffusion")
```




```r
drug_diffusion_raw |> ggplot(aes(x=`Particle Size`, y=`Diffusion Rate`, color=`Surface Charge`)) +
  geom_point() +
  labs(title = "Influence of Particle Size on Drug Delivery Vehicle Diffusion")
```



b) Fit a 2-variable regression model to these data. Do not include any interactions or additional terms in your model. (4 pts)

```
diameter_and_charge_model <- lm(`Diffusion Rate` ~ `Particle Size` + `Surface Charge`,
                                drug_diffusion_raw)
```

c) What are the model coefficients and how do they compare to the prior models' coefficients? Write out all three model equations and evaluate. (6 pts)

```
tidy_diameter_and_charge_model <- tidy(diameter_and_charge_model)

kable(tidy_diameter_and_charge_model)
```

term	estimate	std.error	statistic	p.value
(Intercept)	8.4597	0.0212	399.21	0
`Particle Size`	-0.0203	0.0003	-61.83	0
`Surface Charge`	0.1929	0.0109	17.76	0

```
intercept_diameter_and_charge_model <- pull(filter(tidy_diameter_and_charge_model,
                                                    term=="(Intercept)"), estimate)

size_slope_diameter_and_charge_model <- pull(filter(tidy_diameter_and_charge_model,
                                                     term=="`Particle Size`"), estimate)
```

```r
charge_slope_diameter_and_charge_model <- pull(filter(tidy_diameter_and_charge_model,
                                                      term=="`Surface Charge`"), estimate)


print(paste("Coefficient for intercept: ", intercept_diameter_and_charge_model))
```

```
## [1] "Coefficient for intercept:  8.45965218592573"
```

```r
print(paste("Coefficient for Particle Size: ", size_slope_diameter_and_charge_model))
```

```
## [1] "Coefficient for Particle Size:  -0.0203034215700633"
```

```r
print(paste("Coefficient for Surface Charge: ", charge_slope_diameter_and_charge_model))
```

```
## [1] "Coefficient for Surface Charge:  0.192888719839729"
```

```r
print(paste("Percent change for coefficient for diameter",
            (size_slope_diameter_and_charge_model - slope_size_to_diffusion_model)
            / slope_size_to_diffusion_model * 100))
```

```
## [1] "Percent change for coefficient for diameter -0.341897015314982"
```

```r
print(paste("Percent change for coefficient for charge",
            (charge_slope_diameter_and_charge_model - slope_charge_to_diffusion_model)
            / slope_charge_to_diffusion_model * 100))
```

```
## [1] "Percent change for coefficient for charge -3.99102919523068"
```

Answer: Generic formula for this model $\hat{y} = \beta_0 + \beta_{diameter}x_{diameter} + \beta_{charge}x_{charge}$

Answer: Formula for this model with the estimates substituted $\hat{y} = 8.4597 + -0.0203x_{diameter} + 0.1929x_{charge}$

Answer: Compared to the previous model's the coefficient for particle size and surface charge has remained relatively unchanged with less than a 5% difference.

d) What are $SS_{ParticleSize}$, $SS_{SurfaceCharge}$, $SS_{ParticleSizeY}$, $SS_{SurfaceChargeY}$,SS_{Y}, $SS_{Regression}$, and SS_{Error} (10 pts)?

```r
# Adding lm information to data
augment_diameter_and_charge_model <- augment(diameter_and_charge_model)

# Testing if the regression coefficients are equal to zero
tidy_anova_diameter_and_charge_model <- tidy(car::Anova(diameter_and_charge_model))
```

```r
# Calculating the sum of squares of charge and diffusion
diameter_and_charge_SScharge_y <- pull(summarise(augment_diameter_and_charge_model,
                       SSxy = sum((`Surface Charge` - mean(`Surface Charge`)) *
                                    (`Diffusion Rate` - mean(`Diffusion Rate`)) )
                       )
             )

# Calculating the sum of squares of diameter and diffusion
diameter_and_charge_SSdiameter_y <- pull(summarise(augment_diameter_and_charge_model,
                       SSxy = sum((`Particle Size` - mean(`Particle Size`)) *
                                    (`Diffusion Rate` - mean(`Diffusion Rate`)) )
                       )
             )
```

```r
# Calculating the sum of squares of regression or how well the model represents the data
diameter_and_charge_SSRegression <- pull(summarise(augment_diameter_and_charge_model,
                                SSRegression = sum((.fitted - mean(`Diffusion Rate`)) ^ 2)
                                )
                      )

# Calculating the sum of squares of error/residual
# which measures the discrepancy between the data and the model
diameter_and_charge_SSerror <- pull(summarise(augment_diameter_and_charge_model,
                           SSerror = sum((.fitted - `Diffusion Rate`) ^ 2)
                           )
                 )

kable(tibble(term = c("SSdiameter", "SScharge",
                      "SSdiameter_y", "SScharge_y",
                      "SSy", "SSregression",
                      "SSerror"),
             value = c(SSx_size_to_diffusion , SSx_charge_to_diffusion ,
                       diameter_and_charge_SSdiameter_y, diameter_and_charge_SScharge_y,
                       SSy_diffusion, diameter_and_charge_SSRegression,
                       diameter_and_charge_SSerror))
      )
```

term	value
SSdiameter	89746.0336
SScharge	82.0634
SSdiameter_y	-1828.4028
SScharge_y	16.4871
SSy	40.8932
SSregression	40.3030
SSerror	0.5902

e) What is R^2 for this model (2 pts)?

```r
# Calculating R^2, which is another measure of fit that is between 0 and 1
# Which is the proportion of the variability of Y that can be explained by X
diameter_and_charge_r_squared <-
  (SSy_diffusion - diameter_and_charge_SSerror) / SSy_diffusion

print(paste("R-squared for the model: ", diameter_and_charge_r_squared))
```

```
## [1] "R-squared for the model:  0.985568081925325"
```

f) How does this model's R^2 compare to the prior two models, and why is it so? (4 pts)?

```r
print(paste("Change for R^2 of diameter",
            diameter_and_charge_r_squared - r_squared_charge_to_diffusion))
```

```
## [1] "Change for R^2 of diameter 0.904567370650558"
```

```r
print(paste("Change for R^2 of diameter",
            diameter_and_charge_r_squared - r_squared_size_to_diffusion))
```

```
## [1] "Change for R^2 of diameter 0.0746535596758051"
```

Answer: Compared to the prior two models, the R^2 is higher for a model that uses both particle size and surface charge as the predictor variables than those that use only one or the other. Since we are looking at the same data points with the same diffusion rate across all models, the sum of squares of the dependent variable will stay the same. The change comes from a change in the sum of squares of the residuals. Because the R^2 is increasing, this would be due to a decrease in the sum of squares of the residuals. This indicates that surface charge and particle size explain the variability better than each independently.

g) Use a t-test to evaluate whether each term is zero or non-zero. (4 pts)?

```r
kable(tidy_diameter_and_charge_model)
```

term	estimate	std.error	statistic	p.value
(Intercept)	8.4597	0.0212	399.21	0
`Particle Size`	-0.0203	0.0003	-61.83	0
`Surface Charge`	0.1929	0.0109	17.76	0

```r
diameter_and_charge_tstat_charge <- pull(filter(tidy_diameter_and_charge_model,
                                                 term=="`Surface Charge`"),
                                          statistic)

diameter_and_charge_t_pvalue_charge <- pull(filter(tidy_diameter_and_charge_model,
                                                    term=="`Surface Charge`"),
                                             p.value)

print("t-statistic and associated p-value for slope term of Surface Charge")
```

```
## [1] "t-statistic and associated p-value for slope term of Surface Charge"
```

```r
print(paste("T(", df.residual(diameter_and_charge_model), "):",
            diameter_and_charge_tstat_charge,
            "p = ", diameter_and_charge_t_pvalue_charge))
```

```
## [1] "T( 61 ): 17.763484055445 p =  8.60643701809613e-26"
```

```r
diameter_and_charge_tstat_diameter <- pull(filter(tidy_diameter_and_charge_model,
                                                   term=="`Particle Size`"),
                                            statistic)

diameter_and_charge_t_pvalue_diameter <- pull(filter(tidy_diameter_and_charge_model,
                                                      term=="`Particle Size`"),
                                               p.value)

print("t-statistic and associated p-value for slope term of Particle Size")
```

```
## [1] "t-statistic and associated p-value for slope term of Particle Size"
```

```r
print(paste("T(", df.residual(diameter_and_charge_model), "):",
            diameter_and_charge_tstat_diameter,
            "p = ", diameter_and_charge_t_pvalue_diameter))
```

```
## [1] "T( 61 ): -61.8334315456132 p =  9.75199271855096e-57"
```

Answer: Assuming significance with a p-value less than 0.05, we can reject the null hypothesis that the slope term for the particle size and surface charge is zero.

h) Does size have a significant effect on particle diffusion and, if so, what is that effect? (2 pt)

Answer: After rejecting the null hypothesis that the slope term is zero for particle size, there is evidence that there is a significant linear relationship between the particle size and particle diffusion. This suggests that size has a significant effect on particle diffusion. Because the coefficient for particle size is negative, with an increase in particle size there will be a decrease in particle diffusion.

i) Does surface charge have a significant effect on particle diffusion and, if so, what is that effect? (2 pt)

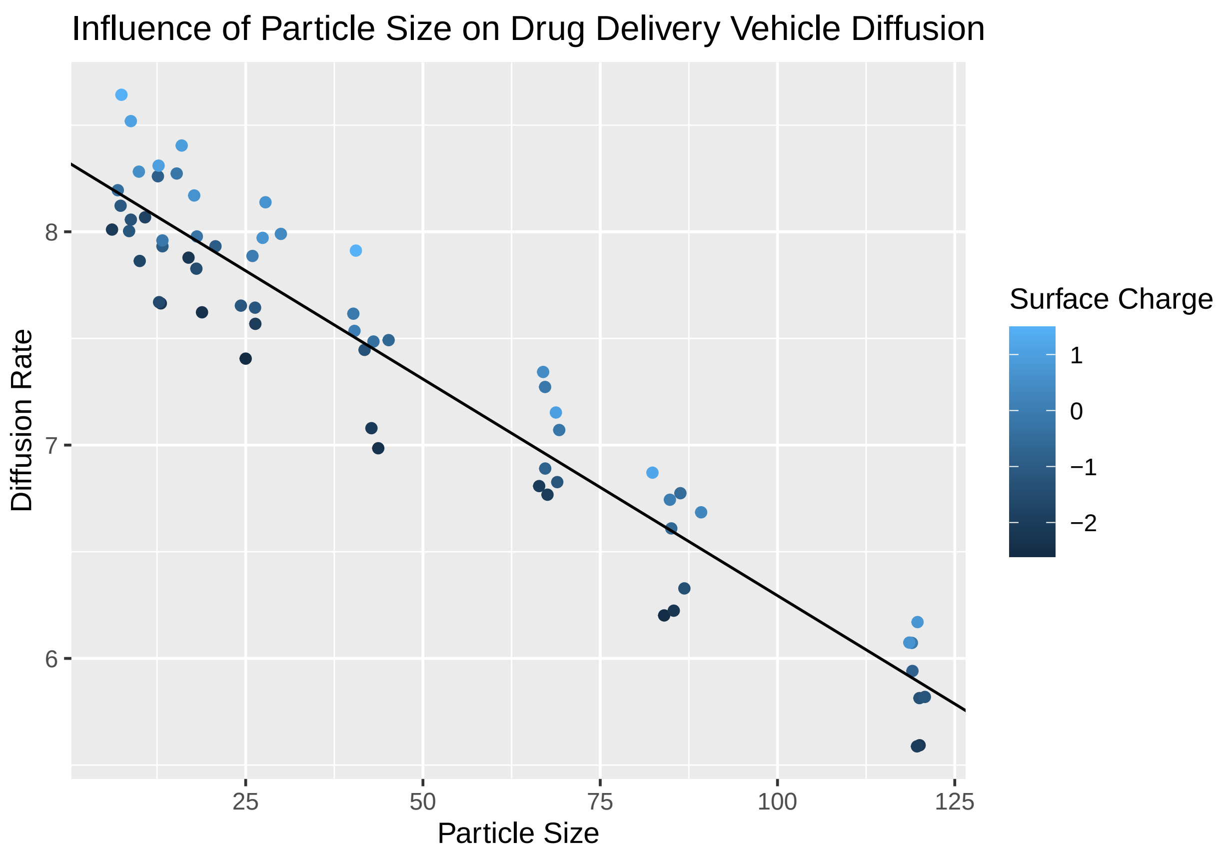
Answer: After rejecting the null hypothesis that the slope term is zero for surface charge, there is evidence that there is a significant linear relationship between the surface charge and particle diffusion. This suggests that charge has a significant effect on particle diffusion. Because the coefficient for surface charge is positive, with an increase in surface charge there will be an increase in particle diffusion.

j) Extend your figures (use your code for your figures from part a of this problem as a base) to show both the raw data and the model using geom_abline. Do not use stat_smooth or other related functions. Think very carefully about where your model's plane is and where/how to represent it in your figures. (8 pt)

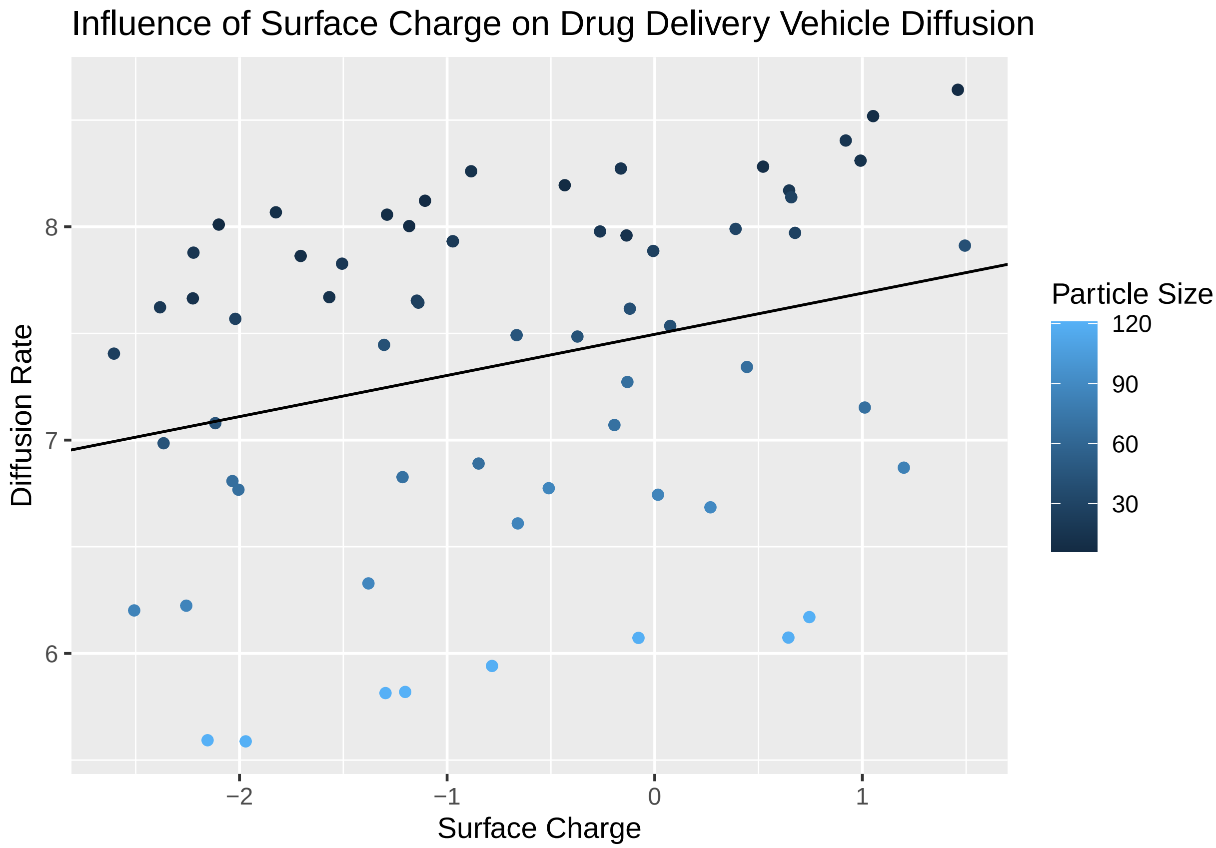
```
mean_charge <- drug_diffusion_raw |> pull(`Surface Charge`) |>
  mean()

mean_size <- drug_diffusion_raw |> pull(`Particle Size`) |>
  mean()


drug_diffusion_raw |> ggplot(aes(x=`Particle Size`, y=`Diffusion Rate`, color=`Surface Charge`)) +
  geom_point() +
  geom_abline(intercept=intercept_diameter_and_charge_model +
                charge_slope_diameter_and_charge_model * mean_charge,
              slope=size_slope_diameter_and_charge_model) +
  labs(title = "Influence of Particle Size on Drug Delivery Vehicle Diffusion")
```




```
drug_diffusion_raw |> ggplot(aes(x=`Surface Charge`, y=`Diffusion Rate`, color=`Particle Size`)) +
  geom_point() +
  geom_abline(intercept=intercept_diameter_and_charge_model +
                size_slope_diameter_and_charge_model * mean_size,
              slope=charge_slope_diameter_and_charge_model ) +
  labs(title = "Influence of Surface Charge on Drug Delivery Vehicle Diffusion")
```



k) If you ran the experiment that resulted in this data, where would you put your focus for optimization of diffusion rate and why? (5 pts)

Answer: Assuming that we want to focus on the variable that has the bigger effect on drug diffusion, there is evidence that particle size should be focused on more. When we look at the P-value of the t-test for both variables in the combined model, they are both below 0.001. But if we look at the R^2, we see that the R^2 is lower for the model using surface charge than the one using particle size. This suggests that particle size has a stronger linear relationship with drug diffusion due to particle size being able to better explain the variance of drug diffusion. Which gives us evidence that particle size has a stronger effect on drug diffusion and should be the focus of further optimization.