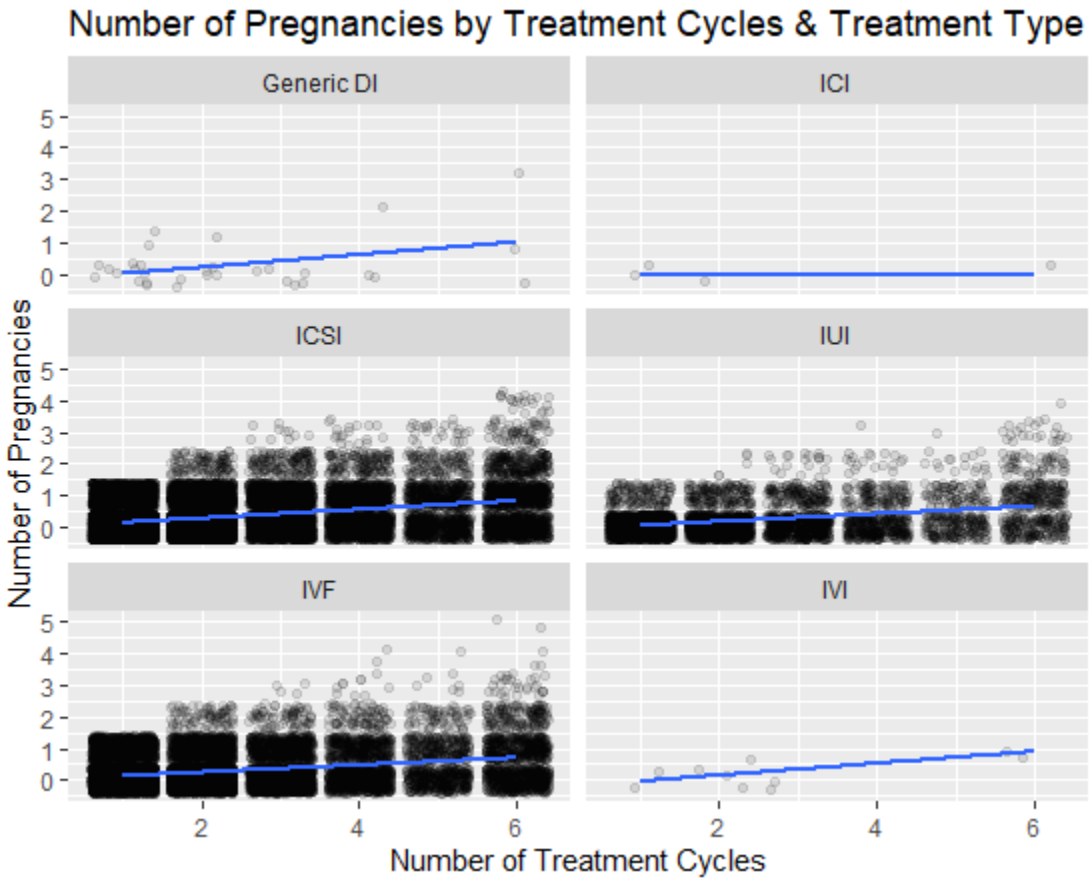
Name: Myra Rust
Date: 24 May 2020
Title: Predicting success rates of IVF

Section 3 – Week 11

What types of plots and tables will help you illustrate the findings to your questions? (Ensure all graph/plots have axis titles, legends if necessary, scales where appropriate, geoms used, etc.).

Research questions

1. What is the success rate of IVF compared to other assisted reproductive procedures?

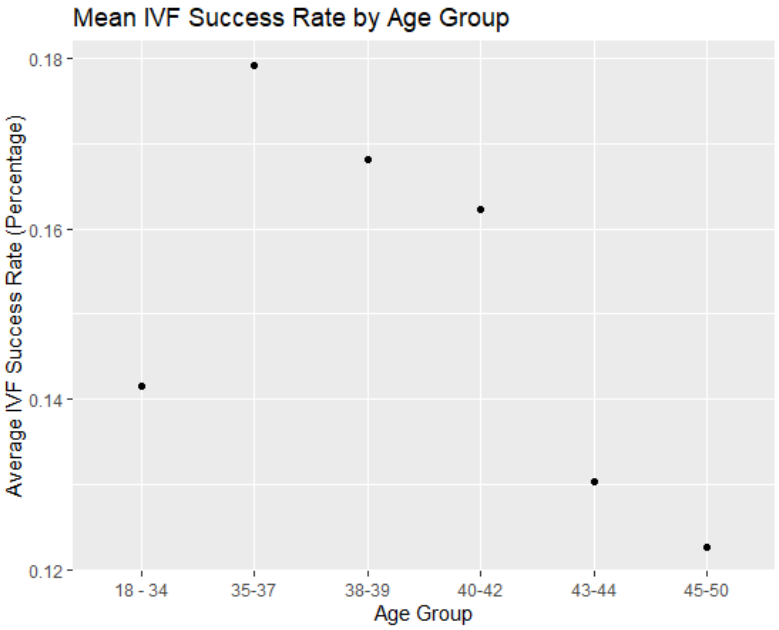



```
Call:
lm(formula = totalPreg ~ totalCycles + specificTreatment, data = ivf3)

Coefficients:
                (Intercept)                totalCycles  specificTreatmentGeneric DI
                    -0.1895                     0.1263                       0.1518
      specificTreatmentICI       specificTreatmentICSI        specificTreatmentIUI
                    -0.1263                     0.2183                       0.1042
      specificTreatmentIVF       specificTreatmentIVI        specificTreatmentUNK
                     0.1822                     0.1439                       0.2864
```

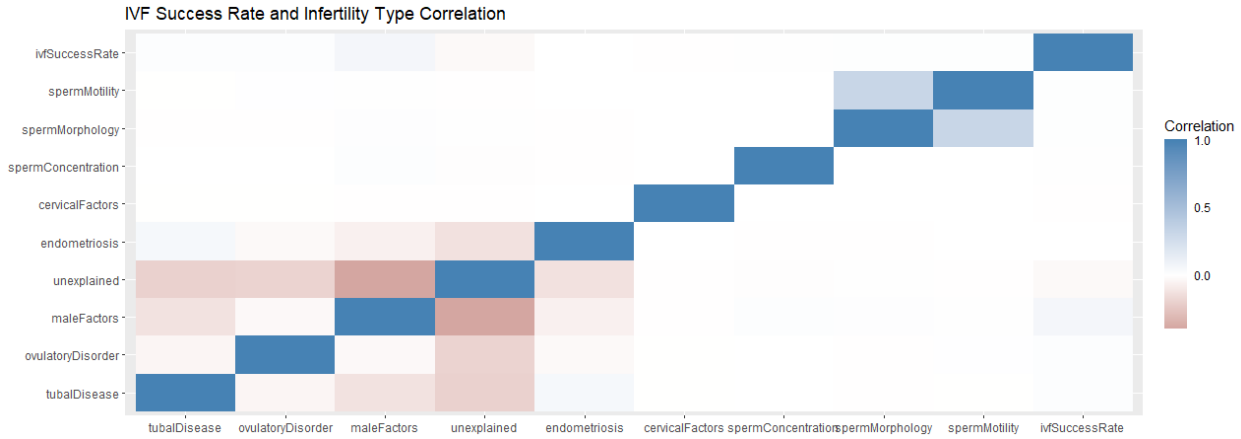
2. What is the success rate of IVF by age?

```
> successByAge
      age meanIVFSuccessRate
1 18 - 34           0.1418549
2   35-37           0.1784737
3   38-39           0.1706650
4   40-42           0.1615033
5   43-44           0.1309622
6   45-50           0.1254985
>
```



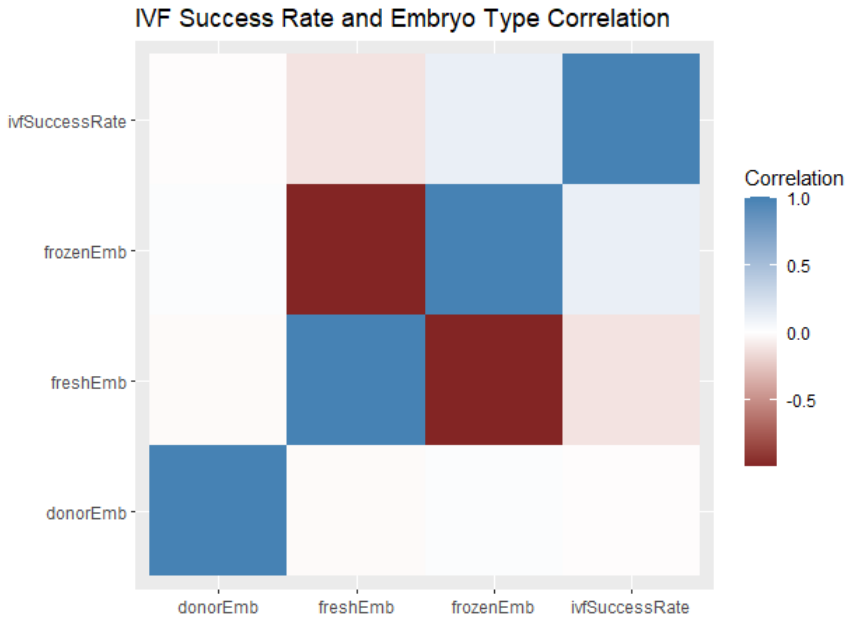
3. What factors have the largest positive/negative effect on the IVF process?
   Correlation between ivfSuccessRate and infertility reasons:

```
> cor(q3_train$ivfSuccessRate, q3_train[, c(2:10)], use = "complete.obs")
     tubalDisease ovulatoryDisorder maleFactors unexplained endometriosis cervicalFactors
[1,]   0.01599547        0.01791701  0.06518915 -0.02832128   0.002389058    -0.002966259
     spermConcentration spermMorphology spermMotility
[1,]         0.00739151      0.00923214   0.008826205
>
```



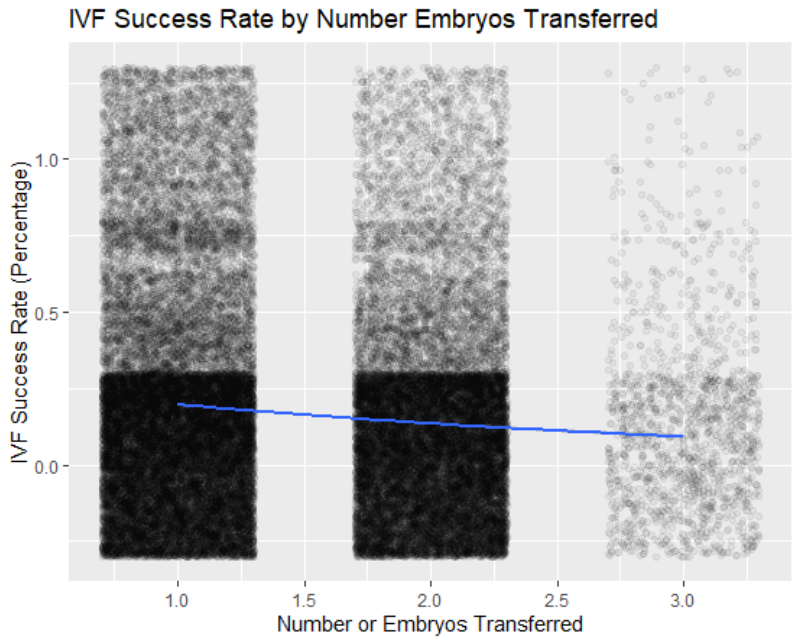
Correlation between ivfSuccessRate and type of embryo used:

```
> #determine correlation between ivfSuccessRate and type of embryo used
> cor(q3_train$ivfSuccessRate, q3_train[, c(11:13)], use = "complete.obs")
          donorEmb   freshEmb frozenEmb
[1,] -0.01212379 -0.1232801 0.1222575
>
```

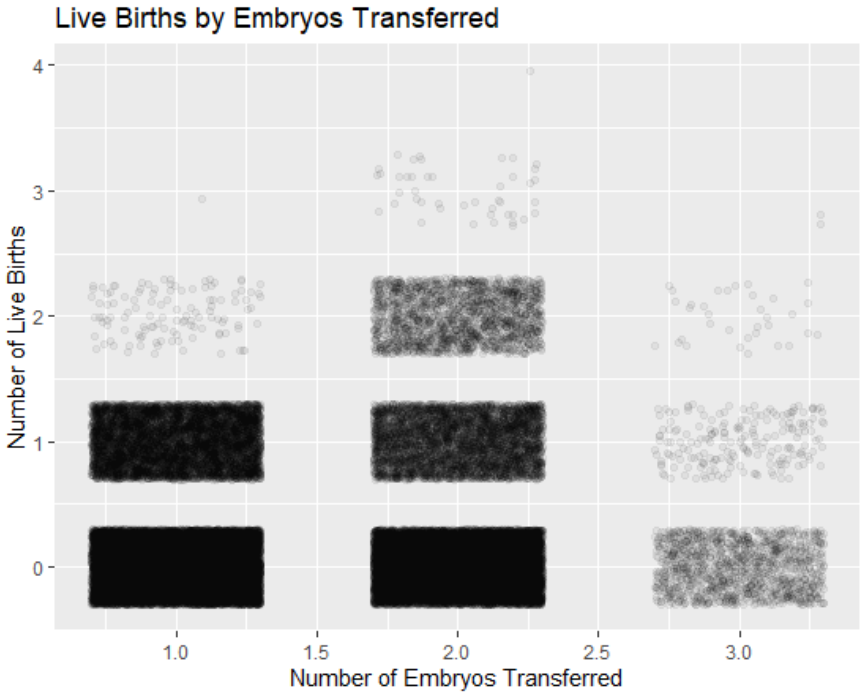


4. Does using multiple embryos increase your chances of success?

```
> successByNumEmb
  numEmbXfer meanIVFSuccessRate
1          1          0.2010181
2          2          0.1305738
3          3          0.1320928
```



5. How much does using multiple embryos increase the likelihood of having twins or multiples?



Number of Embryos Transferred	Likelihood of having multiples
1	1%
2	22%
3	13%

What topics have you learned in this class which most helped you answer your questions?

With no previous experience in statistics or Language R, pretty much everything in this course was vital to being able to answer these questions. Since I learn best and retain the most by doing, I think that the data camp exercises were the best thing for me. However, while working with this data set, I realized a deficiency in my understanding of how to correctly prepare the data for this work. Originally, I thought my data set would be easy to work with, being that it was almost completely binary data. During this last week, I realized that was not true and I wish we had learned more about how to prepare and work with the data when it is not user friendly.

What machine learning techniques do you plan on incorporating to answer your research questions?

I plan to do regression testing to explain the data I have and predict future success rates. Most of my data is not linear, so this will be quite a task using different types of generalized linear models (glms). I currently don't plan to conduct any unsupervised machine learning techniques; I'm planning to focus on regression and not bite of more than I can chew in a week.