Assignment 3:

Emily Gill, Vaiswi Patel, Mehaer Chhabda, Amanda Tsui

Drought Model with interaction term.

```
drought_model_interaction <- glm(Drought.Count ~ delta.temp + Year + delta.temp * Year,
                    family = binomial(link = "logit"),
                    data = NOAAGISSWD)
summary(drought_model_interaction)
```

```
##
## Call:
## glm(formula = Drought.Count ~ delta.temp + Year + delta.temp *
##     Year, family = binomial(link = "logit"), data = NOAAGISSWD)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -164.34308  184.34723  -0.891    0.373
## delta.temp       181.03552  259.47423   0.698    0.485
## Year               0.08222    0.09286   0.885    0.376
## delta.temp:Year   -0.08949    0.12922  -0.693    0.489
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 53.413  on 43  degrees of freedom
## Residual deviance: 46.529  on 40  degrees of freedom
## AIC: 54.529
##
## Number of Fisher Scoring iterations: 4
```

Drought model without interaction term.

```
drought_model<- glm(Drought.Count ~ delta.temp + Year,
                    family = binomial(link = "logit"),
                    data = NOAAGISSWD)
summary(drought_model)
```

```
##
## Call:
## glm(formula = Drought.Count ~ delta.temp + Year, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -87.14359  148.59803  -0.586    0.558
## delta.temp    1.49853    3.78420   0.396    0.692
## Year          0.04364    0.07521   0.580    0.562
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 53.413  on 43  degrees of freedom
## Residual deviance: 46.995  on 41  degrees of freedom
## AIC: 52.995
##
## Number of Fisher Scoring iterations: 4
```

Drought Model with delta.temp only.

Small Model

```
drought_model_delta_temp <- glm(Drought.Count ~ delta.temp,
                     family = binomial(link = "logit"),
                     data = NOAAGISSWD)
summary(drought_model_delta_temp)
```

```
##
## Call:
## glm(formula = Drought.Count ~ delta.temp, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.9172     0.8204  -1.118   0.2636
## delta.temp    3.5189     1.5922   2.210   0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 53.413  on 43  degrees of freedom
## Residual deviance: 47.335  on 42  degrees of freedom
## AIC: 51.335
##
## Number of Fisher Scoring iterations: 4
```

Drought Model with Year only.

Small Model

```
drought_model_Year <- glm(Drought.Count ~ Year,
                     family = binomial(link = "logit"),
                     data = NOAAGISSWD)
summary(drought_model_Year)
```

```
##
## Call:
## glm(formula = Drought.Count ~ Year, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -141.19884   61.85816  -2.283   0.0225 *
## Year           0.07106    0.03097   2.294   0.0218 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 53.413  on 43  degrees of freedom
## Residual deviance: 47.153  on 42  degrees of freedom
## AIC: 51.153
##
## Number of Fisher Scoring iterations: 4
```

Deviance and Akaike's Information Criterion (AIC) values for Drought Model with interaction term.

```
drought_interaction_AIC <- AIC(drought_model_interaction)
drought_interaction_deviance <- deviance(drought_model_interaction)
print(drought_interaction_AIC)
```

```
## [1] 54.52863
```

```
print(drought_interaction_deviance)
```

```
## [1] 46.52863
```

Deviance and Akaike's Information Criterion (AIC) values for Drought Model without interaction term.

```
drought_AIC <- AIC(drought_model)
drought_deviance <- deviance(drought_model)
print(drought_AIC)
```

```
## [1] 52.99482
```

```
print(drought_deviance)
```

```
## [1] 46.99482
```

AIC Comparison:

- Without interaction: 52.99482
- With interaction: 54.52863
- Difference: 54.52863 - 52.99482 = 1.53381 (higher with interaction)
- The lower AIC for the model without interaction indicates it's the better model in terms of balance between fit and complexity

Residual Deviance Comparison:

- Without interaction: 46.99482
- With interaction: 46.52863
- Difference: 46.99482 - 46.52863 = 0.46619 (very small reduction with interaction)
- While the interaction term does reduce the residual deviance slightly, the reduction is minimal (less than 0.5)

Residual Deviance difference of drought model with interaction term and drought model without interaction term.

```
drought_model_difference <- drought_deviance - drought_interaction_deviance
print(drought_model_difference)
```

```
## [1] 0.4661866
```

The model does NOT do better with the interaction term because

1. The reduction in residual deviance (0.46619) when adding the interaction term is very small, suggesting the interaction term adds very little explanatory power
2. The AIC is actually higher (worse) with the interaction term, indicating that the small improvement in fit isn't worth the added complexity
3. Following the principle of parsimony (simpler models are preferred unless there's strong evidence for complexity), we should choose the model without the interaction term

Therefore, based on both metrics, the simpler model without the interaction term is the better choice for modeling drought.

Deviance and Akaike's Information Criterion (AIC) values for Drought Model with delta.temp only.

```
drought_temp_AIC <- AIC(drought_model_delta_temp)
drought_temp_deviance <- deviance(drought_model_delta_temp)
print(drought_temp_AIC)
```

```
## [1] 51.33507
```

```
print(drought_temp_deviance)
```

```
## [1] 47.33507
```

Delta temp-only model:

- AIC = 51.33507

Model with both year and delta temp (from previous):

- AIC = 52.99482

AIC Comparison:

- Delta temp only: 51.33507
- Year and delta temp: 52.99482
- Difference: 52.99482 - 51.33507 = 1.65975 (lower/better for temp-only)
- Deviance comparison: 0.34025

Analysis: The temperature-only model has a lower AIC (51.33507) compared to the combined model (52.99482), indicating that the temperature-only model provides a better balance of fit and complexity. The difference of about 1.66 suggests that adding the year variable doesn't improve the model enough to justify the added complexity.

Residual Deviance Difference of Drought Model with delta.temp only (small model) and Drought Model without interaction term.

```
drought_temp_difference <- drought_temp_deviance - drought_deviance
print(drought_temp_difference)
```

```
## [1] 0.3402506
```

The model does better with just delta temp because:

1. The AIC is lower for the temperature-only model (51.33507 vs 52.99482)

2. While the residual deviance is slightly lower with the combined model, the reduction (0.34025) is very small
3. The increase in AIC for the combined model suggests that adding the year variable adds complexity without sufficient improvement in model fit
4. Following the principle of parsimony, the simpler temperature-only model is preferred since the more complex model doesn't provide substantial improvement

Therefore, based on both AIC and residual deviance analysis, the temperature-only model is the better choice for modeling drought in this case.

Deviance and Akaike's Information Criterion (AIC) values for Drought Model with Year only.

```
drought_year_AIC <- AIC(drought_model_Year)
drought_year_deviance <- deviance(drought_model_Year)
print(drought_year_AIC)
```

```
## [1] 51.15299
```

```
print(drought_year_deviance)
```

```
## [1] 47.15299
```

Year-only model:
- AIC = 51.15299
- Residual Deviance = 47.15299

Year + delta temp model (combined):
- AIC = 52.99482
- Residual Deviance = 46.99482

Analysis:
1. AIC Comparison:
    - Year-only: 51.15299
    - Combined model: 52.99482
    - Difference: 52.99482 - 51.15299 = 1.84183 (higher/worse for combined)
2. Residual Deviance Comparison:
    - Year-only: 47.15299
    - Combined model: 46.99482
    - Difference: 47.15299 - 46.99482 = 0.15817

Residual Deviance Difference of Drought Model with Year only (small model) and Drought Model without interaction term.

```
drought_year_difference <- drought_year_deviance - drought_deviance
print(drought_year_difference)
```

```
## [1] 0.1581774
```

The model does better with the year-only model because:

1. The residual deviance difference (0.15817) is very small, indicating that adding temperature to the year model provides minimal improvement in fit
2. The AIC is notably lower for the year-only model (51.15299 vs 52.99482)

3. The increase in AIC (1.84183) for the combined model suggests that adding the temperature variable adds complexity without meaningful improvement in model fit
4. Following the principle of parsimony, the simpler year-only model is preferred since adding temperature doesn't provide substantial improvement

Final Analysis:

The year-only model appears to be the best choice for modeling drought because:

- It has the lowest AIC of all models
- Adding complexity (temperature or interaction terms) increases AIC without substantial improvement in residual deviance
- It provides the best balance of model fit and simplicity

Wildfire Model with interaction term.

```
wildfire_model_interaction <- glm(Wildfire.Count ~ delta.temp + Year + delta.temp * Year,
                    family = binomial(link = "logit"),
                    data = NOAAGISSWD)
summary(wildfire_model_interaction)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ delta.temp + Year + delta.temp *
##     Year, family = binomial(link = "logit"), data = NOAAGISSWD)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -188.74474  218.72821  -0.863    0.388
## delta.temp       -52.78261  306.29915  -0.172    0.863
## Year               0.09409    0.11004   0.855    0.393
## delta.temp:Year    0.02670    0.15263   0.175    0.861
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 43.717  on 40  degrees of freedom
## AIC: 51.717
##
## Number of Fisher Scoring iterations: 4
```

Wildfire model without interaction term.

```
wildfire_model <- glm(Wildfire.Count ~ delta.temp + Year,
                      family = binomial(link = "logit"),
                      data = NOAAGISSWD)
summary(wildfire_model)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ delta.temp + Year, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -212.80370  169.82370  -1.253    0.210
## delta.temp     0.81269    4.12303   0.197    0.844
## Year           0.10610    0.08587   1.236    0.217
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 43.748  on 41  degrees of freedom
## AIC: 49.748
##
## Number of Fisher Scoring iterations: 4
```

Wildfire Model with delta.temp only.

Small Model

```
wildfire_model_delta_temp <- glm(Wildfire.Count ~ delta.temp,
                      family = binomial(link = "logit"),
                      data = NOAAGISSWD)
summary(wildfire_model_delta_temp)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ delta.temp, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.118      1.033  -3.020  0.00253 **
## delta.temp     5.699      1.799   3.168  0.00154 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 45.406  on 42  degrees of freedom
## AIC: 49.406
##
## Number of Fisher Scoring iterations: 4
```

Wildfire Model with Year only.

Small Model

```
wildfire_model_Year <- glm(Wildfire.Count ~ Year,
                     family = binomial(link = "logit"),
                     data = NOAAGISSWD)
summary(wildfire_model_Year)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ Year, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -243.44788   72.16830  -3.373 0.000743 ***
## Year           0.12163    0.03606   3.373 0.000743 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 43.787  on 42  degrees of freedom
## AIC: 47.787
##
## Number of Fisher Scoring iterations: 4
```

Deviance and Akaike's Information Criterion (AIC) values for Wildfire Model with interaction term.

```
wildfire_interaction_AIC <- AIC(wildfire_model_interaction)
wildfire_interaction_deviance <- deviance(wildfire_model_interaction)
print(wildfire_interaction_AIC)
```

```
## [1] 51.71748
```

```
print(wildfire_interaction_deviance)
```

```
## [1] 43.71748
```

Deviance and Akaike's Information Criterion (AIC) values for Wildfire Model without interaction term.

```
wildfire_AIC <- AIC(wildfire_model)
wildfire_deviance <- deviance(wildfire_model)
print(wildfire_deviance)
```

```
## [1] 43.74829
```

```
print(wildfire_AIC)
```

```
## [1] 49.74829
```

- AIC Comparison
    - Model with Interaction Term AIC: 51.71748
    - Model without Interaction Term AIC: 49.74829
    - Difference: 51.71748 - 49.74829 = 1.969
    - Analysis: The model without the interaction term is more favorable based on its lower AIC value. The lower AIC value is preferential because it balances the fit and complexity of the model. Overall, this indicates that the addition of an interaction term introduces complexity without enough improvement in model fit.
- Residual Deviance Comparison
    - Model with Interaction Term: 43.71748
    - Model without Interaction Term: 43.74829
    - Difference: 43.74829 - 43.71748 = 0.031
    - Analysis: The model without the interaction term is better at balancing the fit and simplicity of the model, based on the AIC. While the interaction model does provide a slightly lower residual deviance, it has an extra complexity associated with its higher AIC.

Residual Deviance Difference of Wildfire Model with interaction term and Wildfire Model without interaction term.

```
wildfire_model_difference <- wildfire_deviance - wildfire_interaction_deviance
print(wildfire_model_difference)
```

```
## [1] 0.03080962
```

- The residual deviance between the wildfire models with and without the interaction terms is 0.03080962, which is a very small value. This small difference may indicate that both models fit data similarly and the inclusion of an interaction term does not significantly impact the model's ability to explain the variance in the data. Furthermore, given the lower AIC value for the model without the interaction term suggests that it may be a better overall choice due to its better model fit and simplicity.

Deviance and Akaike's Information Criterion (AIC) values for Wildfire Model with delta.temp only.

```
wildfire_temp_AIC <- AIC(wildfire_model_delta_temp)
wildfire_temp_deviance <- deviance(wildfire_model_delta_temp)
print(wildfire_temp_AIC)
```

```
## [1] 49.40551
```

```
print(wildfire_temp_deviance)
```

```
## [1] 45.40551
```

AIC Comparisons of Small Models:

- Temp-only model AIC: 49.40551
- Year-only model AIC: 47.787
- The year only model has a lower AIC compared to the temperature only model, thus indicating that the year performs better in explaining the data, as well as balancing model and fit complexity compared to temperature alone.

AIC and residual deviance of delta temp only with the model with both year and delta temp analysis:

- AIC of delta temp only: 49.40551
- Residual deviance of delta temp only: 45.40551
- AIC of year + delta temp: 49.74829
- Residual deviance of year + delta temp: 43.74829
- Analysis: Based on the AIC, the delta temp only model has a slightly lower AIC compared to the model with both delta temp and year. This indicates that the delta temp model is better balanced in terms of the model complexity and fit. However, the model with both year and delta temp has a lower residual deviance, which indicates that the model fits the data better, indicating that the addition of the year leads to a small improvement in the model fit.

Residual Deviance Difference of Wildfire Model with delta.temp only (small model) and Wildfire Model without interaction term.

```
wildfire_temp_difference <- wildfire_temp_deviance - wildfire_deviance
print(wildfire_temp_difference)
```

```
## [1] 1.657221
```

Based on the residual deviance difference of 1.657221 shown in the output, the model performs better with the year variable added (without delta.temp) compared to just delta.temp alone. The positive difference indicates that the delta.temp-only model has higher residual deviance, meaning it explains less of the variance in the data. This suggests that year is a more important predictor variable for the wildfire model than temperature alone.

This is consistent with what we see in the AIC values as well - the year-only model has an AIC of 47.787 compared to the delta.temp-only model's AIC of 49.40551, indicating better model performance with just the year variable.

Deviance and Akaike's Information Criterion (AIC) values for Wildfire Model with Year only.

```
wildfire_year_AIC <- AIC(wildfire_model_Year)
wildfire_year_deviance <- deviance(wildfire_model_Year)
print(wildfire_year_AIC)
```

```
## [1] 47.78727
```

```
print(wildfire_year_deviance)
```

```
## [1] 43.78727
```

- AIC and residual deviance of year only with the model with both year and delta temp analysis:

Year-only model:

- AIC: 47.79
- Residual deviance: 43.79

Model with both year and delta temp:

- AIC: 49.74829
- Residual deviance: 43.74829

The year-only model performs better with a lower AIC (47.79 vs 49.74829), suggesting it's a better fit. The residual deviances are very similar (43.79 vs 43.75), indicating that adding delta temp doesn't meaningfully improve model fit. Therefore, the simpler year-only model is preferred.

Residual Deviance Difference of Wildfire Model with Year only (small model) with Wildfire Model without interaction term.

```
wildfire_year_difference <- wildfire_year_deviance - wildfire_deviance
print(wildfire_year_difference)
```

```
## [1] 0.03898426
```

- Year-only model and year + temp model using residual difference analysis:

Based on the residual deviance difference of 0.03898426, which is very small, there is minimal difference between the year-only model and the year + temp model. Since this difference is positive but very close to zero, the year + temp model performs only marginally better than the year-only model. Given such a small improvement, the simpler year-only model would be preferred since it doesn't add any unnecessary complexity.

Final analysis: The year and delta temp model without interaction is the best model for portraying wildfires due to the fact that:

- It has a low residual deviance.
- It has the lowest AIC.
- Combines the temporal based trends captured in Year with the climate-related factors captured in Delta Temp which contribute to a better understanding and accuracy of the model
- Maintains a balance between balance and fit