OceanNow_survey_analysis

Dr. Liam D. Bailey

21/01/2022

Introduction

Data preparation

Load cleaned data

```
clean_data <- read.csv(here::here("data/clean_data.csv"))
```

Remove records with unknown gender (6).

```
table(clean_data$Gender)
```

```
##
## Female   Male  Other
##     49     47      6
```

```
clean_data <- clean_data %>%
  dplyr::filter(Gender != "Other")
```

Use all data from Berlin. For Kiel, use only one survey from each individual (avoid paired data).

```
## NEED TO SET SEED TO MAKE SURE WE GET THE SAME DATA FROM KIEL EACH TIME!
set.seed(123)

Berlin_data <- clean_data %>%
  filter(Location == "BERLIN")

Kiel_data <- clean_data %>%
  filter(Location == "KIEL") %>%
  group_by(PersonalCode) %>%
  #Randomly select one survey from each individual
  dplyr::slice(sample(1:n(), 1))

clean_data <- bind_rows(Berlin_data, Kiel_data)

table(clean_data$Location)
```
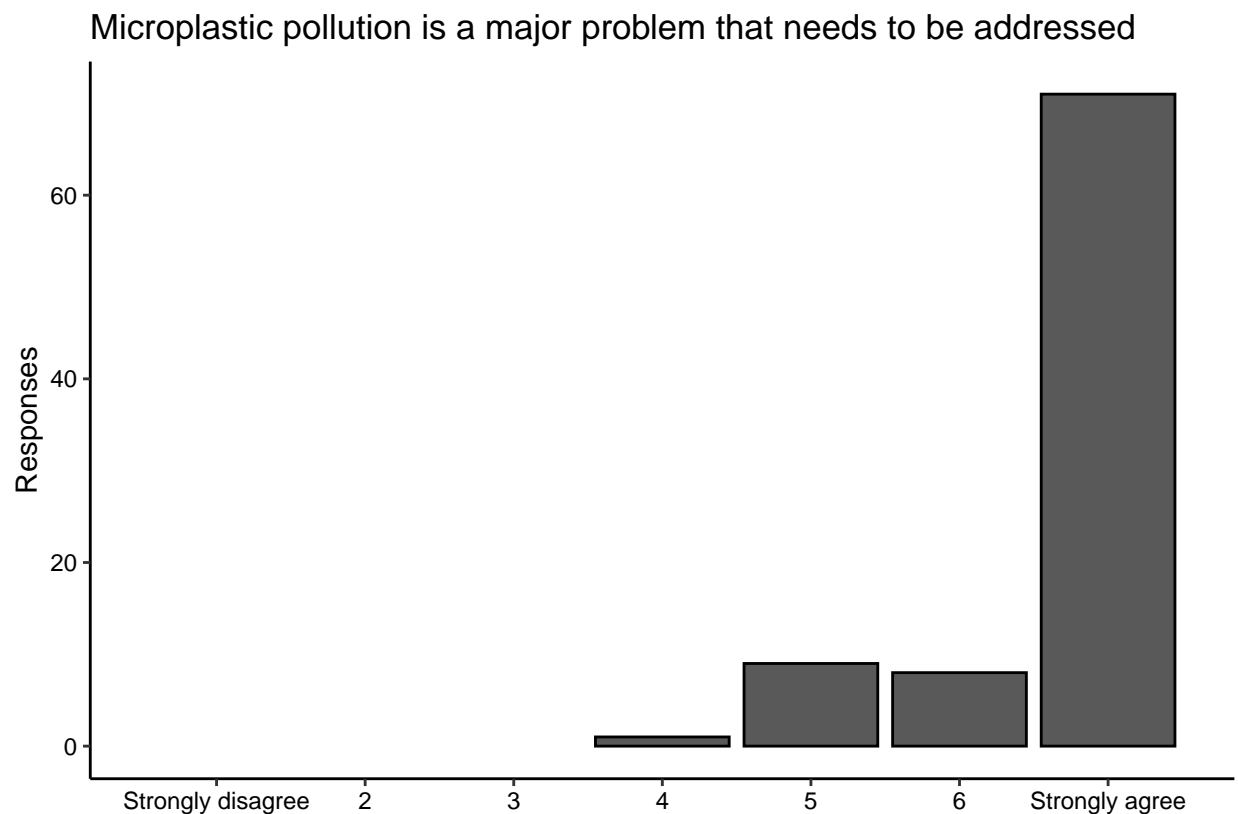
```
##
## BERLIN   KIEL
##     75     14
```

Analysis

1. Microplastic pollution is a major problem that needs to be addressed

Step 1: Check whether data can be grouped for logistic regression.

```
ggplot() +
  geom_bar(data = clean_data, aes(x = Microplastic_importance_LK), colour = "black") +
  scale_x_continuous(name = "", limits = c(0.5, 7.5),
                     breaks = seq(1, 7, 1), labels = c("Strongly disagree", 2:6, "Strongly agree")) +
  scale_y_continuous(name = "Responses") +
  labs(title = "Microplastic pollution is a major problem that needs to be addressed") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```


There is a clear bias towards 'strongly agree' so we will analyse this data using a logistic regression (7 v. <7).

Step 2: Fit preliminary model

We include an interaction between gender and treatment to see if effect of exhibition has different effects on men and women. To start with we use age groups defined by 'generations'. We will check to see if this differs if we use age groups that are more balanced. We include Location to make sure there is no difference between Kiel and Berlin.

```
##Turn Likert scale into binary variable
Q1_data <- clean_data %>%
  dplyr::mutate(response_category = dplyr::case_when(Microplastic_importance_LK == 7 ~ 1,
                                                    Microplastic_importance_LK != 7 ~ 0))

##Create model
Q1_model <- glm(response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location,
                data = Q1_data, family = binomial(link = "logit"))

summary(Q1_model)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Gen_Age_group + Location, family = binomial(link = "logit"),
##     data = Q1_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4232   0.3508   0.4971   0.7639   1.0815
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 2.5528     1.1017   2.317   0.0205 *
## GenderMale                 -1.5944     0.8809  -1.810   0.0703 .
## Gen_Age_groupGen_Z         -0.7286     1.0176  -0.716   0.4740
## Gen_Age_groupMillenials    -0.7222     0.8853  -0.816   0.4146
## LocationKIEL                0.1240     1.1201   0.111   0.9118
## GenderFemale:TreatmentTRUE  0.2045     0.9846   0.208   0.8355
## GenderMale:TreatmentTRUE    0.8409     0.7264   1.158   0.2470
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 89.623  on 88  degrees of freedom
## Residual deviance: 81.340  on 82  degrees of freedom
## AIC: 95.34
##
## Number of Fisher Scoring iterations: 4
```
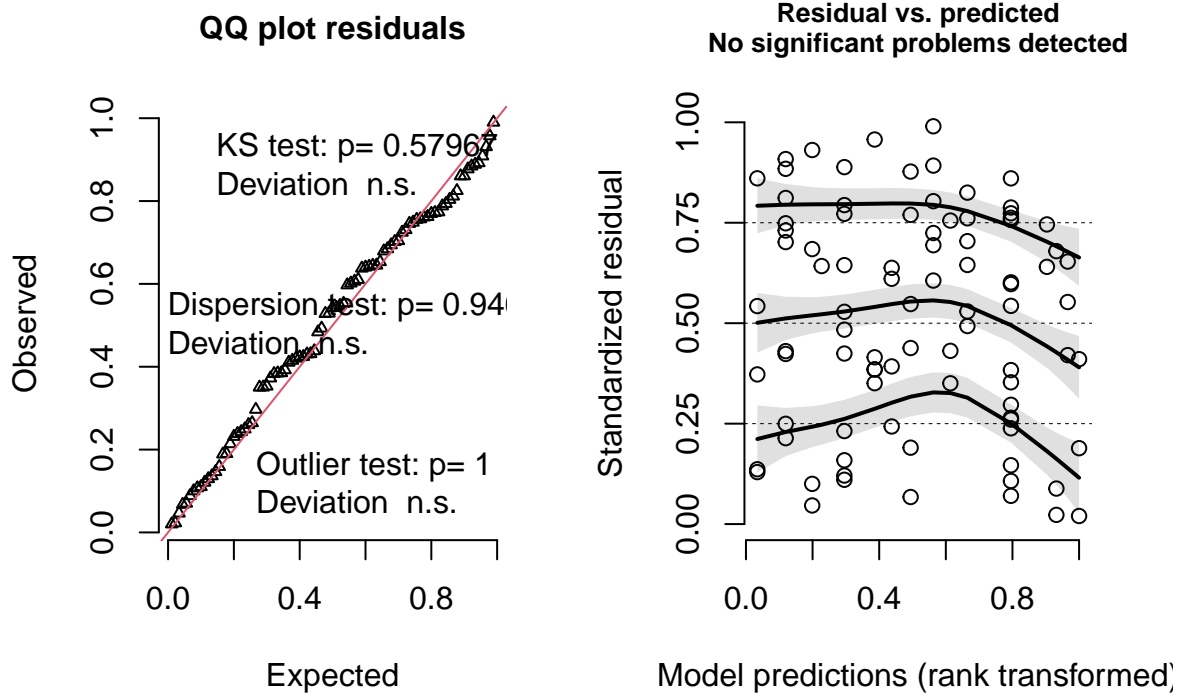
No significant results. Results suggest that males are less likely to answer strongly agree than women. Treatment has no significant effect, although effect size is much larger in men than women.
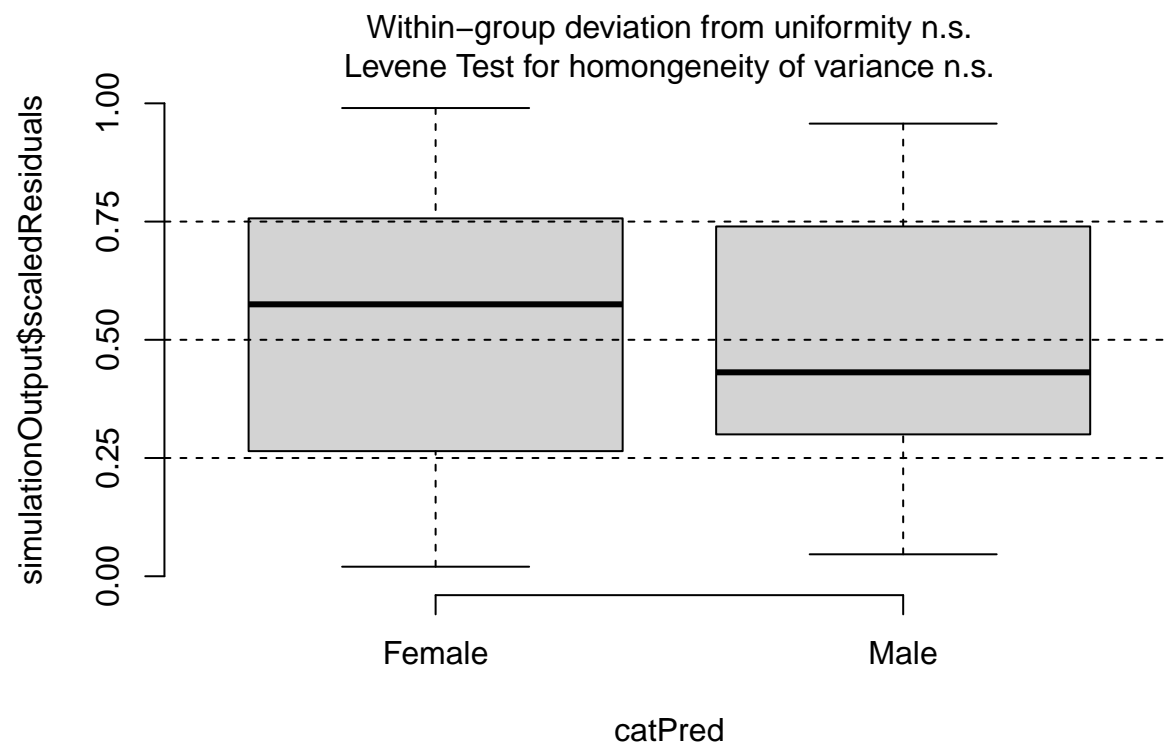
Step 3: Check model reliability

Use DHARMa to check for issues of heteroskedasticity or outliers.

```
##Check model using DHARMa
simulationOutput <- simulateResiduals(Q1_model, seed = 123, n = 5000)
plot(simulationOutput)
```

DHARMa residual diagnostics

QQ plot residuals

KS test: p= 0.5796?
Deviation n.s.

Dispersion test: p= 0.94
Deviation n.s.

Outlier test: p= 1
Deviation n.s.

Residual vs. predicted
No significant problems detected
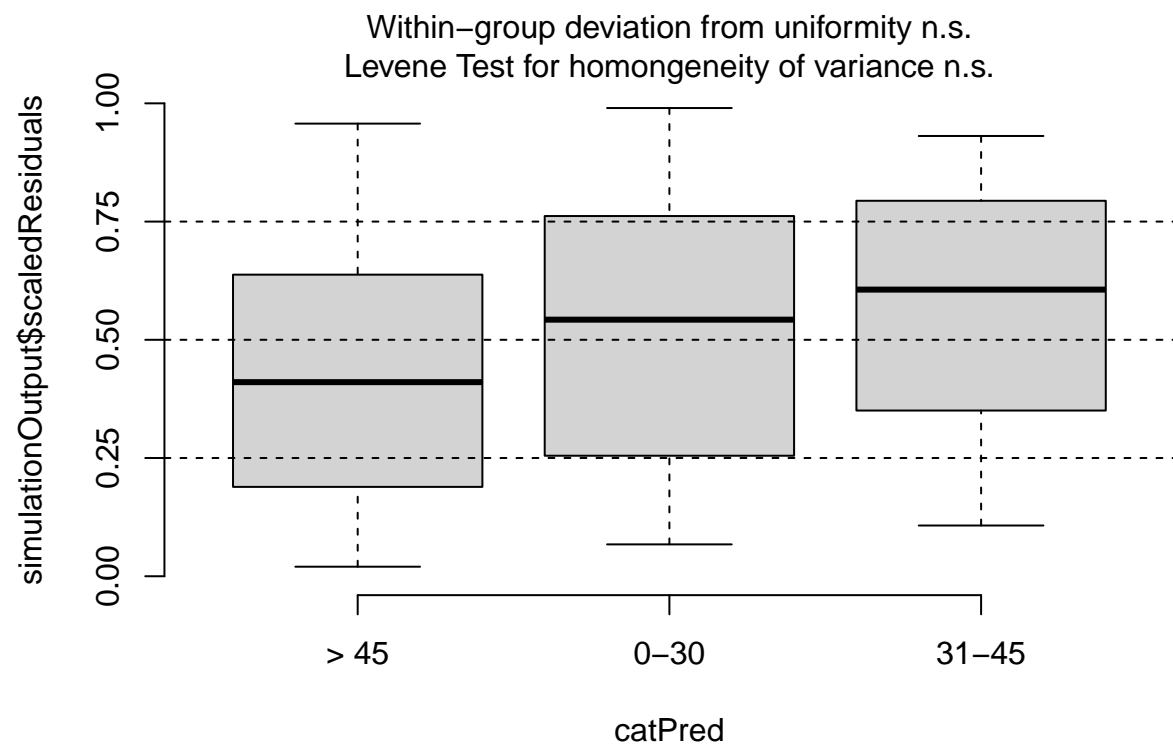


```
#Check there are no unusual residual patterns for each predictor
plotResiduals(simulationOutput, form = as.factor(Q1_data$Gender))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.

```
plotResiduals(simulationOutput, form = as.factor(Q1_data$Age_group))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.

```
plotResiduals(simulationOutput, form = as.factor(Q1_data$Treatment))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
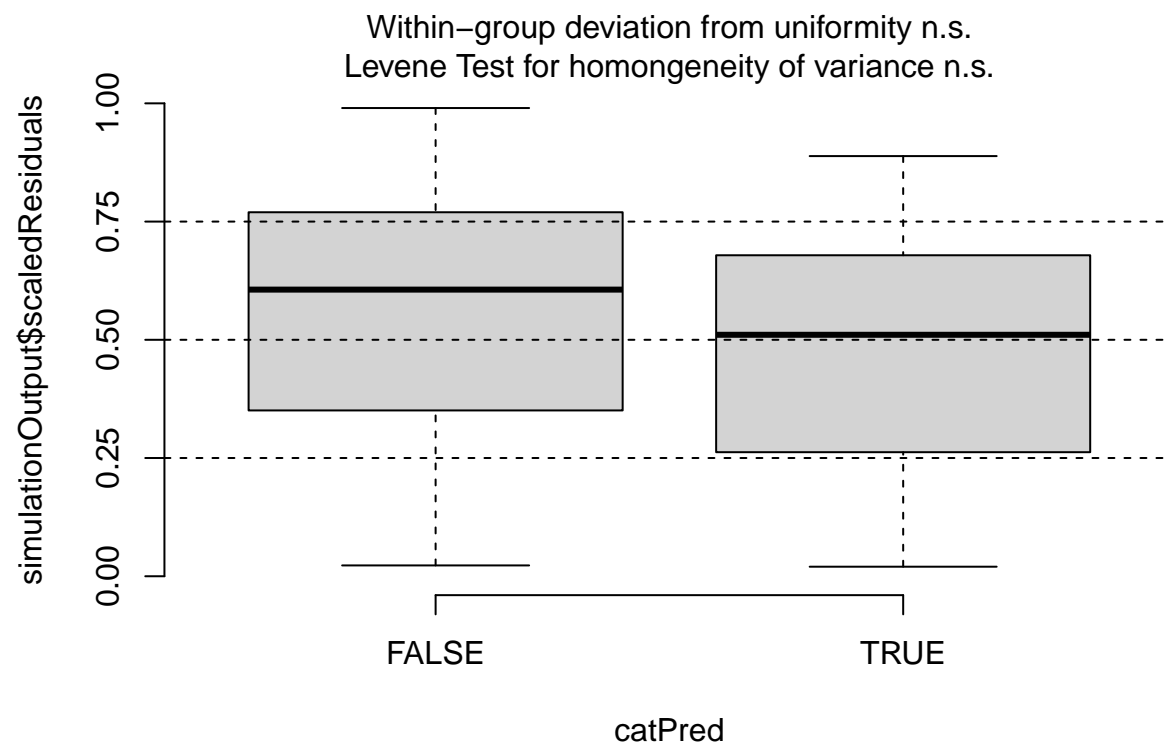
```
plotResiduals(simulationOutput, form = as.factor(Q1_data$Location))
```
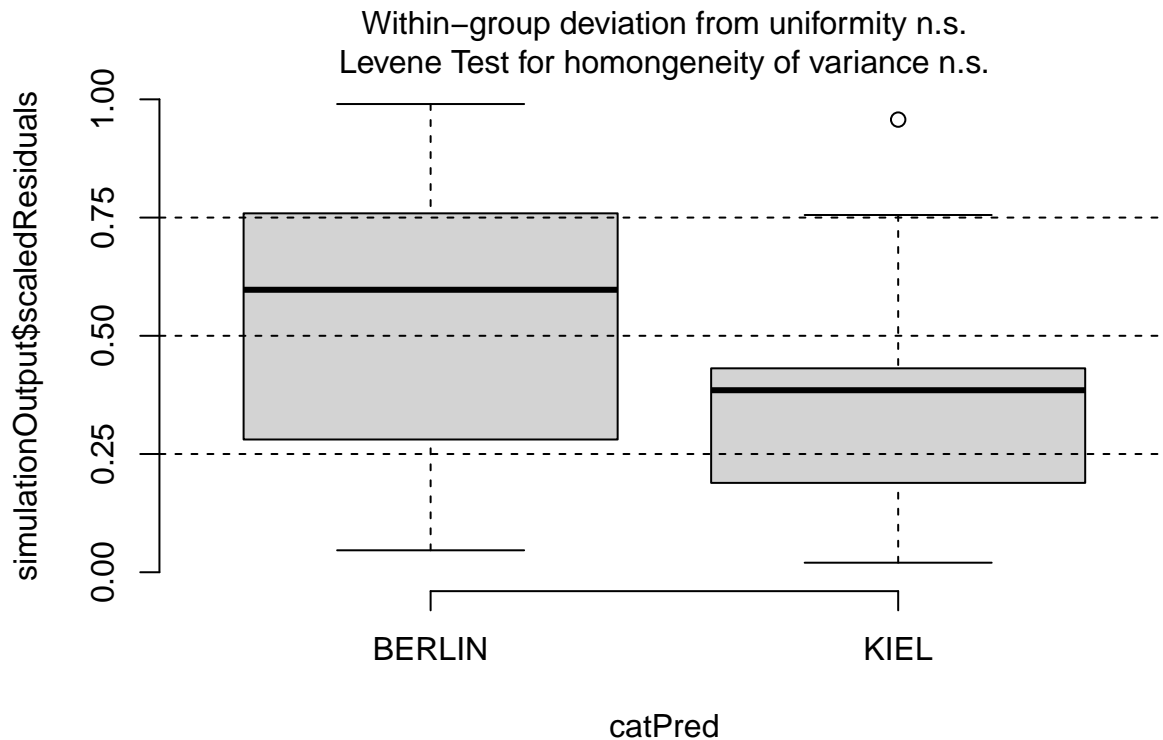
No apparent issues with model.

Step 3b: Check that fitting model with alternative age group gives the same results.

```
##Create model
Q1_model_age <- glm(response_category ~ Gender + Gender:Treatment + Age_group + Location,
                    data = Q1_data, family = binomial(link = "logit"))

summary(Q1_model_age)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Age_group + Location, family = binomial(link = "logit"),
##     data = Q1_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3767   0.3715   0.5387   0.6289   1.1000
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.9477     1.0783   1.806   0.0709 .
## GenderMale                -1.5333     0.8760  -1.750   0.0800 .
```

8

```
## Age_group0-30                    -0.2296     0.8694  -0.264    0.7917
## Age_group31-45                     0.5522     0.9473   0.583    0.5599
## LocationKIEL                       0.6765     1.0524   0.643    0.5204
## GenderFemale:TreatmentTRUE         0.1389     0.9773   0.142    0.8870
## GenderMale:TreatmentTRUE           0.5536     0.7645   0.724    0.4690
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 89.623  on 88  degrees of freedom
## Residual deviance: 80.888  on 82  degrees of freedom
## AIC: 94.888
##
## Number of Fisher Scoring iterations: 4
```

Results are qualitatively the same. We can continue just using 'generational' cut-offs. However, it would be good to identify age trends in different responses to see if there is a more appropriate age grouping.

```r
Q1_data %>%
  mutate(age_cutoff = cut(Age, breaks = c(seq(15, 45, 5), 85)),
         age_cutoff_numeric = as.numeric(age_cutoff)) %>%
  group_by(age_cutoff_numeric) %>%
  summarise(binom::binom.wilson(x = sum(response_category), n = n()),
            age_cutoff = first(age_cutoff),
            n = n()) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = age_cutoff_numeric, ymin = lower, ymax = upper), width = 0.25) +
      geom_point(aes(x = age_cutoff_numeric, y = mean), size = 3, shape = 21, fill = "grey45") +
      geom_text(aes(x = age_cutoff_numeric + 0.15, y = mean, label = n), size = 4) +
      scale_x_continuous(breaks = seq(1:max(.$age_cutoff_numeric)),
                         labels = .$age_cutoff) +
      scale_y_continuous(limits = c(0, 1)) +
      theme_classic()}
```

Could we even analyse this as a continous variable with quadratic effect?

```
##Create model
Q1_model_agequad <- glm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q1_data, family = binomial(link = "logit"))

summary(Q1_model_agequad)
```
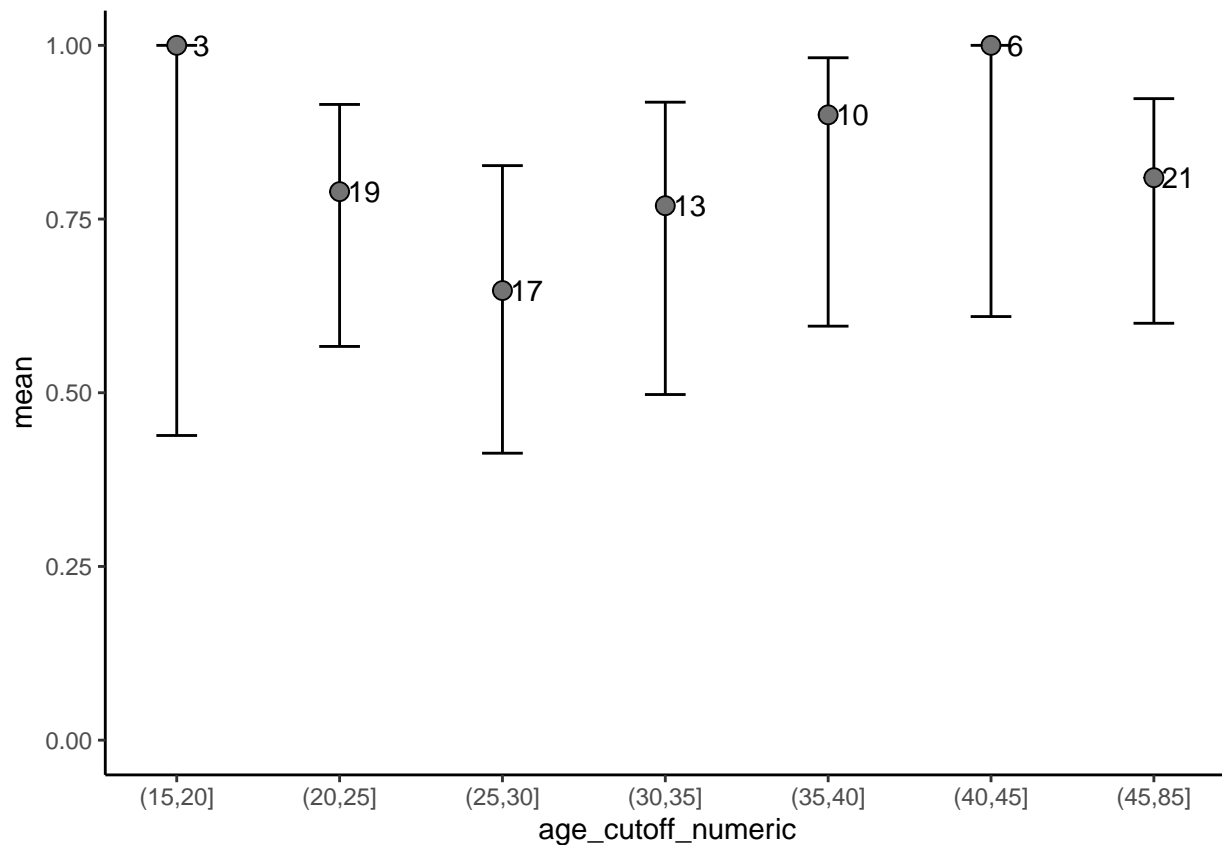
```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     poly(Age, 2) + Location, family = binomial(link = "logit"),
##     data = Q1_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4615   0.3643   0.5131   0.6852   1.0906
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)               2.0352     0.8005   2.542   0.0110 *
## GenderMale               -1.5838     0.8860  -1.788   0.0739 .
## poly(Age, 2)1             2.1040     4.1985   0.501   0.6163
## poly(Age, 2)2            -0.0439     3.1764  -0.014   0.9890
## LocationKIEL              0.3142     1.1783   0.267   0.7897
## GenderFemale:TreatmentTRUE 0.1384    0.9811   0.141   0.8878
```

```
## GenderMale:TreatmentTRUE      0.8014      0.7310     1.096    0.2730
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 89.623  on 88  degrees of freedom
## Residual deviance: 81.825  on 82  degrees of freedom
## AIC: 95.825
##
## Number of Fisher Scoring iterations: 5
```

```
AIC(Q1_model, Q1_model_age, Q1_model_agequad)
```

```
##                  df      AIC
## Q1_model          7 95.34035
## Q1_model_age      7 94.88830
## Q1_model_agequad  7 95.82513
```

Model with age groups is (slightly) better than one using generational cut-off.

Step 4: Plot results

```
## Plot results (gender and treatment)
plot_data <- Q1_data %>%
  dplyr::group_by(Gender, Treatment) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(response_category), n = n()), .groups = "drop")

ggplot() +
  geom_errorbar(data = plot_data, aes(x = Treatment, ymin = lower, ymax = upper, group = Gender),
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data, aes(x = Treatment, y = mean, fill = Gender),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "Microplastic pollution is a major problem that needs to be addressed") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

Microplastic pollution is a major problem that needs to be addressed
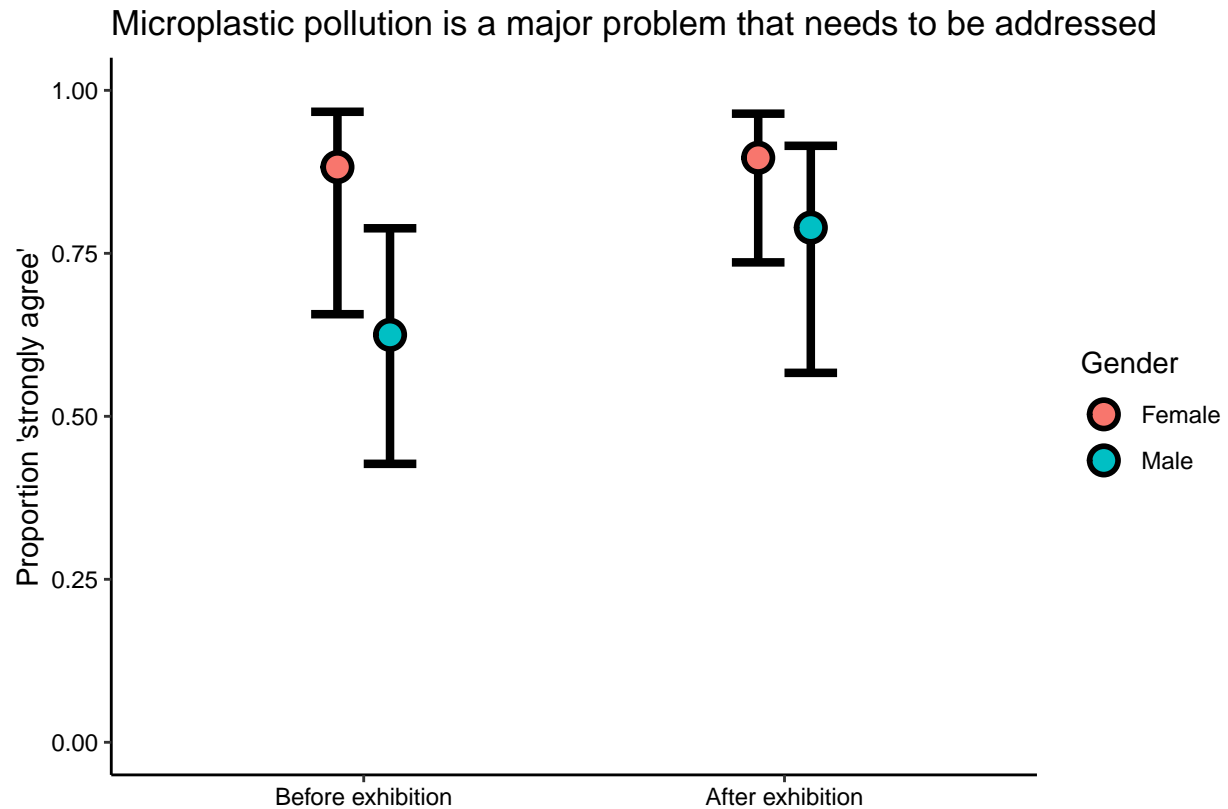
2. As humans we are able to act against microplastic pollution

Step 1: Check whether data can be grouped for logistic regression.

```
ggplot() +
  geom_bar(data = clean_data, aes(x = Viable_action_LK), colour = "black") +
  scale_x_continuous(name = "", limits = c(0.5, 7.5),
                     breaks = seq(1, 7, 1), labels = c("Strongly disagree", 2:6, "Strongly agree")) +
  scale_y_continuous(name = "Responses") +
  labs(title = "As humans we are able to act against microplastic pollution") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

As humans we are able to act against microplastic pollution

There is a clear bias towards 'strongly agree' so we will analyse this data using a logistic regression (7 v. <7).

Step 2: Fit preliminary model

We include an interaction between gender and treatment to see if effect of exhibition has different effects on men and women. To start with we use age groups defined by 'generations'. We will check to see if this differs if we use age groups that are more balanced. We include Location to make sure there is no difference between Kiel and Berlin.

```r
##Turn Likert scale into binary variable
Q2_data <- clean_data %>%
  dplyr::mutate(response_category = dplyr::case_when(Viable_action_LK == 7 ~ 1,
                                                     Viable_action_LK != 7 ~ 0))


##Create model
Q2_model <- glm(response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location,
                data = Q2_data, family = binomial(link = "logit"))

summary(Q2_model)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##      Gen_Age_group + Location, family = binomial(link = "logit"),
```

```
##      data = Q2_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0490  -0.7425   0.5113   0.7563   1.6871
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  1.9703     0.9324   2.113   0.0346 *
## GenderMale                  -1.5297     0.7668  -1.995   0.0460 *
## Gen_Age_groupGen_Z          -1.5881     0.8906  -1.783   0.0746 .
## Gen_Age_groupMillenials     -0.1707     0.7854  -0.217   0.8280
## LocationKIEL                -0.2876     0.9272  -0.310   0.7564
## GenderFemale:TreatmentTRUE   0.1690     0.7956   0.212   0.8318
## GenderMale:TreatmentTRUE     0.8354     0.6828   1.223   0.2211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 107.521  on 88  degrees of freedom
## Residual deviance:  94.892  on 82  degrees of freedom
## AIC: 108.89
##
## Number of Fisher Scoring iterations: 4
```

Males are significantly less likely to answer strongly agree than women. Treatment has no significant effect, although effect size is much larger in men than women. Suggestion that Gen Z are less likely to answer strongly agree than GenX+.

Step 3: Check model reliability

Use DHARMa to check for issues of heteroskedasticity or outliers.

```
##Check model using DHARMa
simulationOutput <- simulateResiduals(Q2_model, seed = 123, n = 5000)
plot(simulationOutput)
```
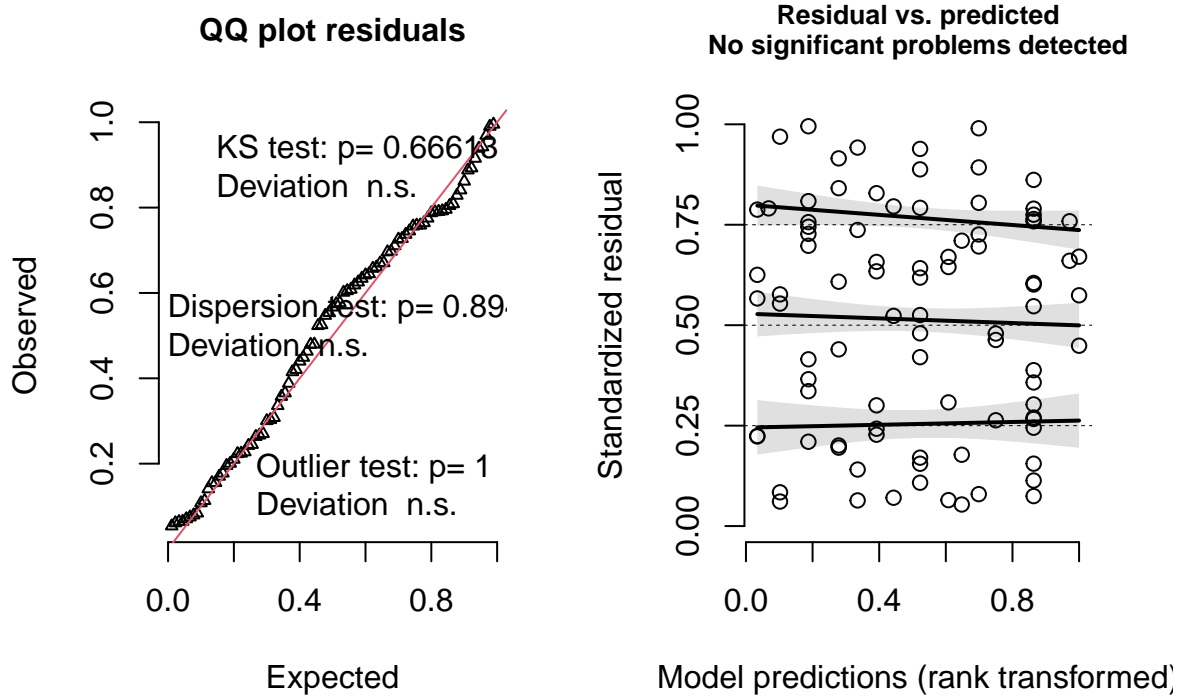
DHARMa residual diagnostics

QQ plot residuals

KS test: p= 0.6661
Deviation n.s.

Dispersion test: p= 0.894
Deviation n.s.

Outlier test: p= 1
Deviation n.s.

Observed

Expected

Residual vs. predicted
No significant problems detected

Standardized residual

Model predictions (rank transformed)

```
#Check there are no unusual residual patterns for each predictor
plotResiduals(simulationOutput, form = as.factor(Q2_data$Gender))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
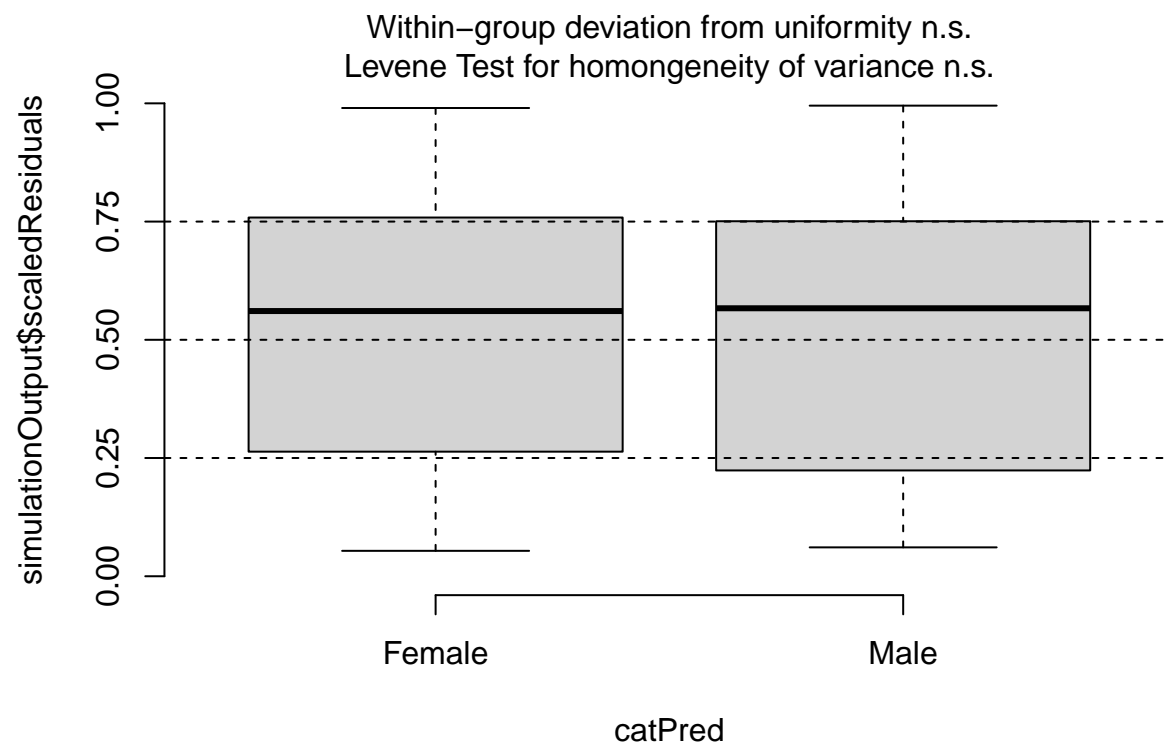
```
plotResiduals(simulationOutput, form = as.factor(Q2_data$Age_group))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
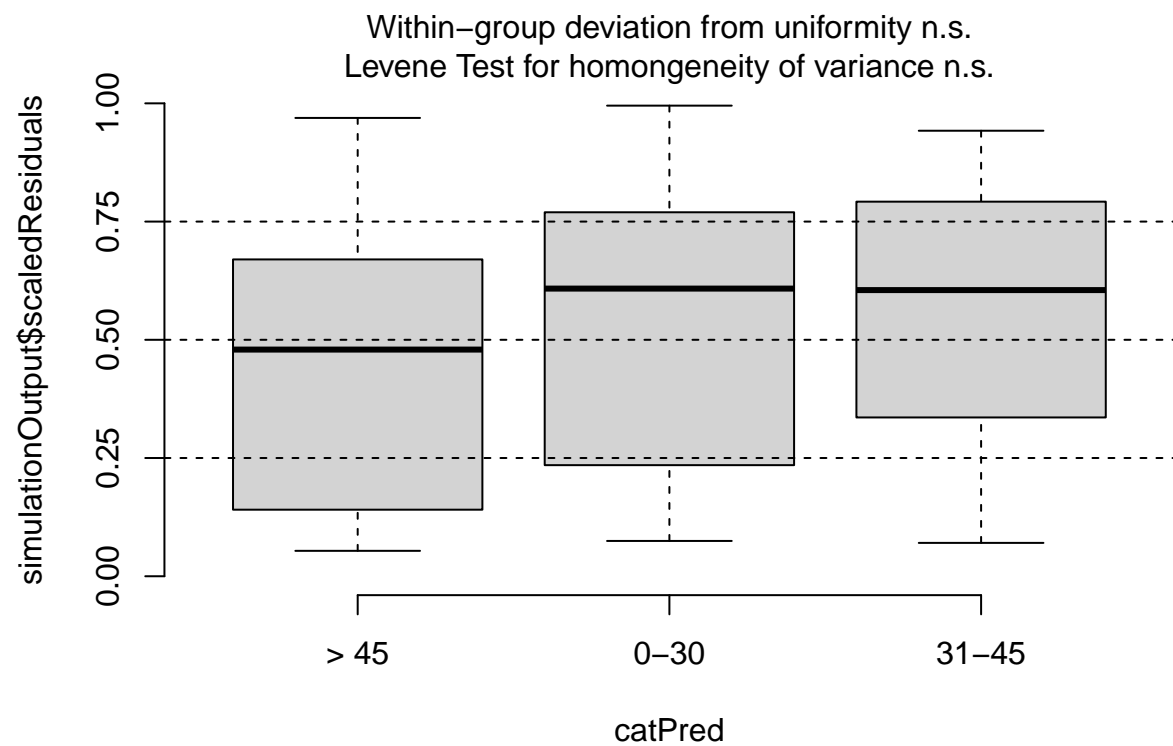
```
plotResiduals(simulationOutput, form = as.factor(Q2_data$Treatment))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
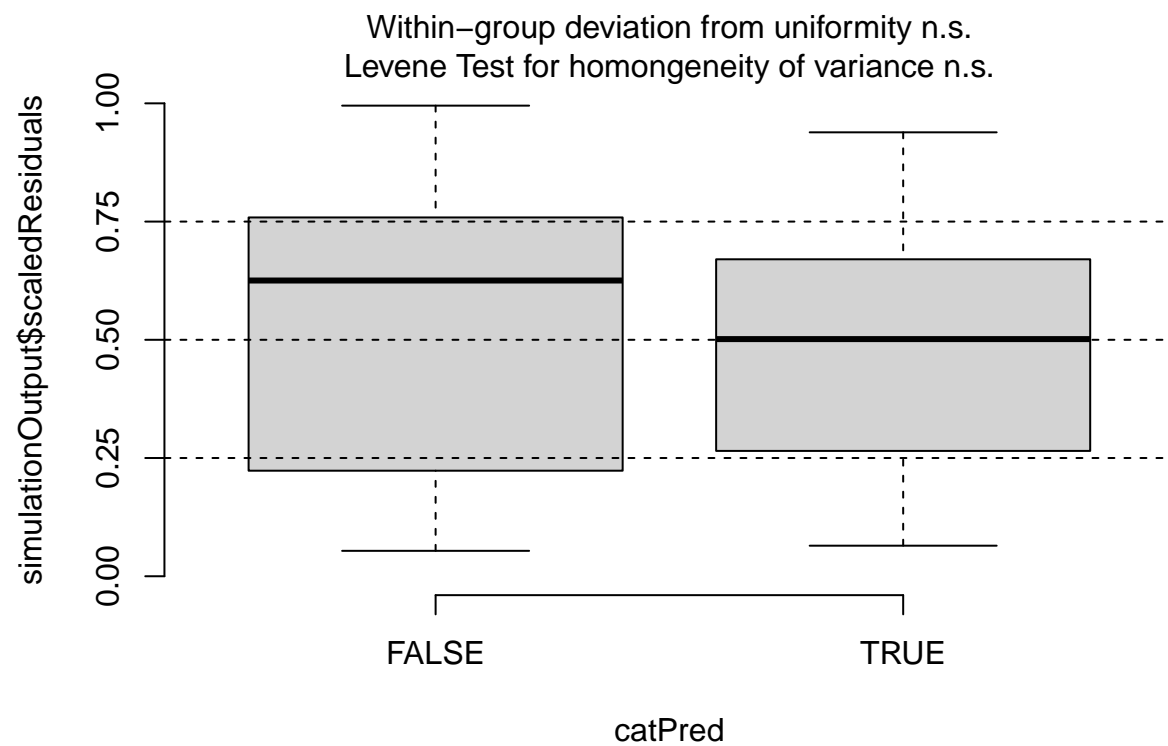
```
plotResiduals(simulationOutput, form = as.factor(Q2_data$Location))
```
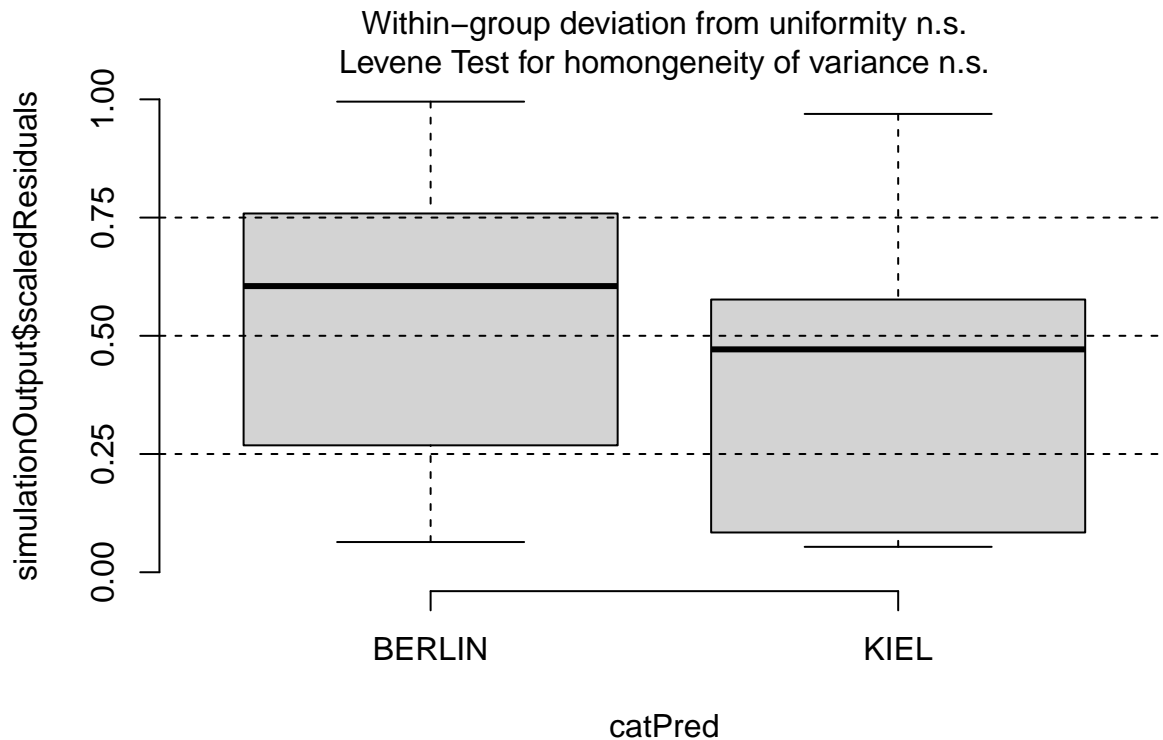
No apparent issues with model.

Step 3b: Check that fitting model with alternative age group gives the same results.

```r
##Create model
Q2_model_age <- glm(response_category ~ Gender + Gender:Treatment + Age_group + Location,
                    data = Q2_data, family = binomial(link = "logit"))

summary(Q2_model_age)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Age_group + Location, family = binomial(link = "logit"),
##     data = Q2_data)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.0623  -1.0570   0.6608   0.7784   1.3027
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               1.44385    0.91335   1.581   0.1139
## GenderMale               -1.20564    0.71348  -1.690   0.0911 .
```

```
## Age_group0-30                 -0.52809     0.80504  -0.656   0.5118
## Age_group31-45                 0.19919     0.85069   0.234   0.8149
## LocationKIEL                   -0.03328     0.88512  -0.038   0.9700
## GenderFemale:TreatmentTRUE   0.35665     0.76218   0.468   0.6398
## GenderMale:TreatmentTRUE     0.80056     0.71319   1.123   0.2616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 107.521  on 88  degrees of freedom
## Residual deviance:  98.804  on 82  degrees of freedom
## AIC: 112.8
##
## Number of Fisher Scoring iterations: 4
```

Results are less strong when using other age categories. Why is this? What is the underlying pattern of age in the raw data?

```
Q2_data %>%
  mutate(age_cutoff = cut(Age, breaks = c(seq(15, 45, 5), 85)),
         age_cutoff_numeric = as.numeric(age_cutoff)) %>%
  group_by(age_cutoff_numeric) %>%
  summarise(binom::binom.wilson(x = sum(response_category), n = n()),
            age_cutoff = first(age_cutoff)) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = age_cutoff_numeric, ymin = lower, ymax = upper), width = 0.25) +
      geom_point(aes(x = age_cutoff_numeric, y = mean), size = 3, shape = 21, fill = "grey45") +
      scale_x_continuous(breaks = seq(1:max(.$age_cutoff_numeric)),
                         labels = .$age_cutoff) +
      scale_y_continuous(limits = c(0, 1)) +
      theme_classic()}
```
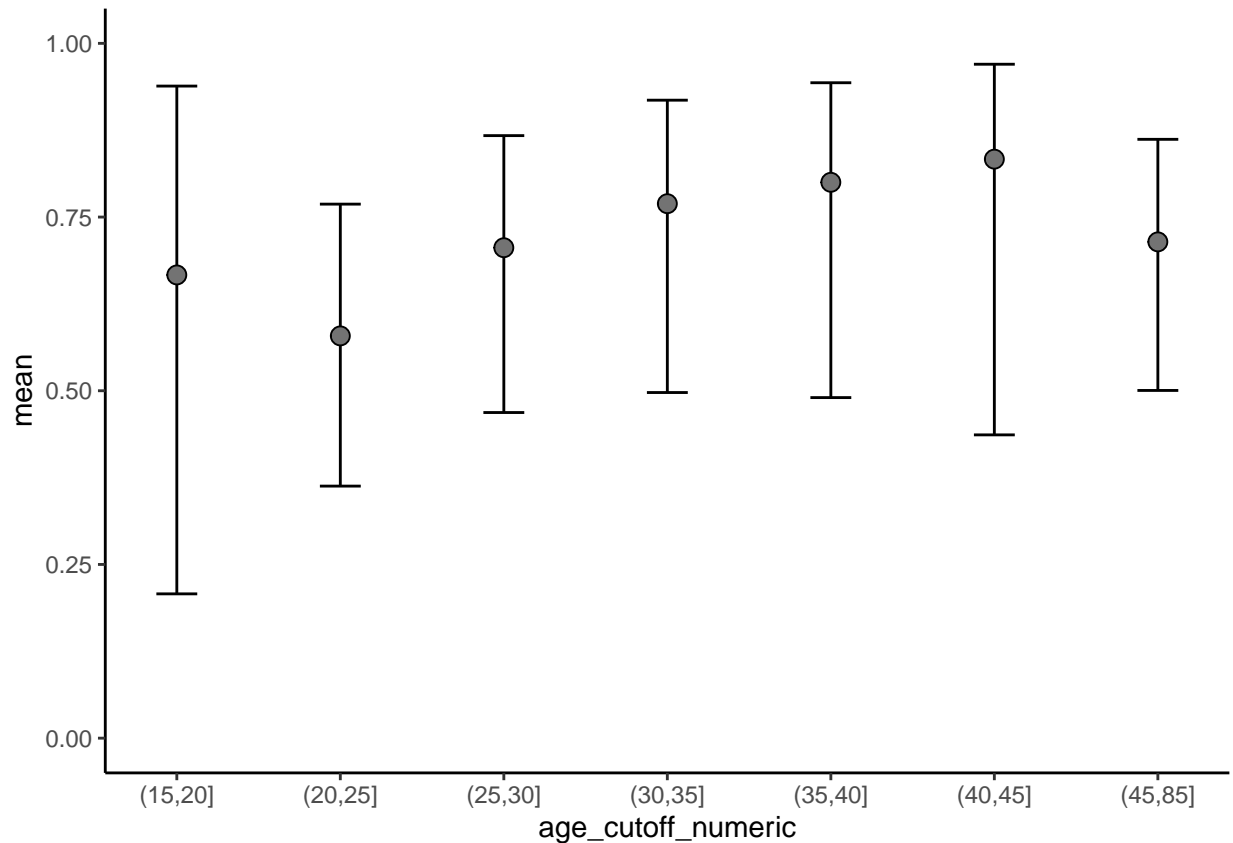
When using generations, 15-25 is its own group which have more pessimistic views that middle aged people. When using balanced age groups, Gen Z is grouped with some middle age and the effect is diluted. Could we even analyse this as a continous variable with quadratic effect?

```
##Create model
Q2_model_agequad <- glm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q2_data, family = binomial(link = "logit"))

summary(Q2_model_agequad)


##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     poly(Age, 2) + Location, family = binomial(link = "logit"),
##     data = Q2_data)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.9440  -0.9792   0.6196   0.7871   1.3932
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.5871     0.6525   2.432   0.0150 *
## GenderMale                -1.3225     0.7429  -1.780   0.0750 .
## poly(Age, 2)1              8.3595     4.5146   1.852   0.0641 .
## poly(Age, 2)2              0.9017     3.0572   0.295   0.7680
```

```
## LocationKIEL                    -1.4047      1.0942  -1.284    0.1992
## GenderFemale:TreatmentTRUE    0.3138      0.7750   0.405    0.6855
## GenderMale:TreatmentTRUE      0.9984      0.7031   1.420    0.1556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 107.521  on 88  degrees of freedom
## Residual deviance:  95.778  on 82  degrees of freedom
## AIC: 109.78
##
## Number of Fisher Scoring iterations: 5


AIC(Q2_model, Q2_model_age, Q2_model_agequad)


##                    df      AIC
## Q2_model            7 108.8921
## Q2_model_age        7 112.8038
## Q2_model_agequad    7 109.7778
```

Using generational cut-off is best fit comparing AIC.

Step 4: Plot results


```
## Plot results (gender and treatment)
plot_data_treatment <- Q2_data %>%
  dplyr::group_by(Gender, Treatment) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(response_category), n = n()), .groups = "drop")

treatment_effect <- ggplot() +
  geom_errorbar(data = plot_data_treatment, aes(x = Treatment, ymin = lower, ymax = upper, group = Gend
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data_treatment, aes(x = Treatment, y = mean, fill = Gender),
            shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
      title = "As humans we are able to act against microplastic pollution",
      subtitle = "Effect of gender") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))

plot_data_age <- Q2_data %>%
  group_by(Gen_Age_group, Treatment) %>%
  summarise(binom::binom.wilson(x = sum(response_category), n = n()), .groups = "drop") %>%
  mutate(Gen_Age_group = ordered(Gen_Age_group, levels = c("Gen_Z", "Millenials", "Gen_X_Bommer")))

age_effect <- ggplot() +
  geom_errorbar(data = plot_data_age, aes(x = Treatment, ymin = lower, ymax = upper, group = Gen_Age_gr
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
```

```
    geom_point(data = plot_data_age, aes(x = Treatment, y = mean, fill = Gen_Age_group),
              shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
    scale_y_continuous(limits = c(0, 1)) +
    scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
    scale_fill_manual(values = c("#33658a", "#4c9f70", "#ca3c25")) +
    labs(y = "Proportion 'strongly agree'", x = "",
         subtitle = "Effect of age") +
    theme_classic() +
    theme(axis.text = element_text(colour = "black"))

treatment_effect + age_effect
```
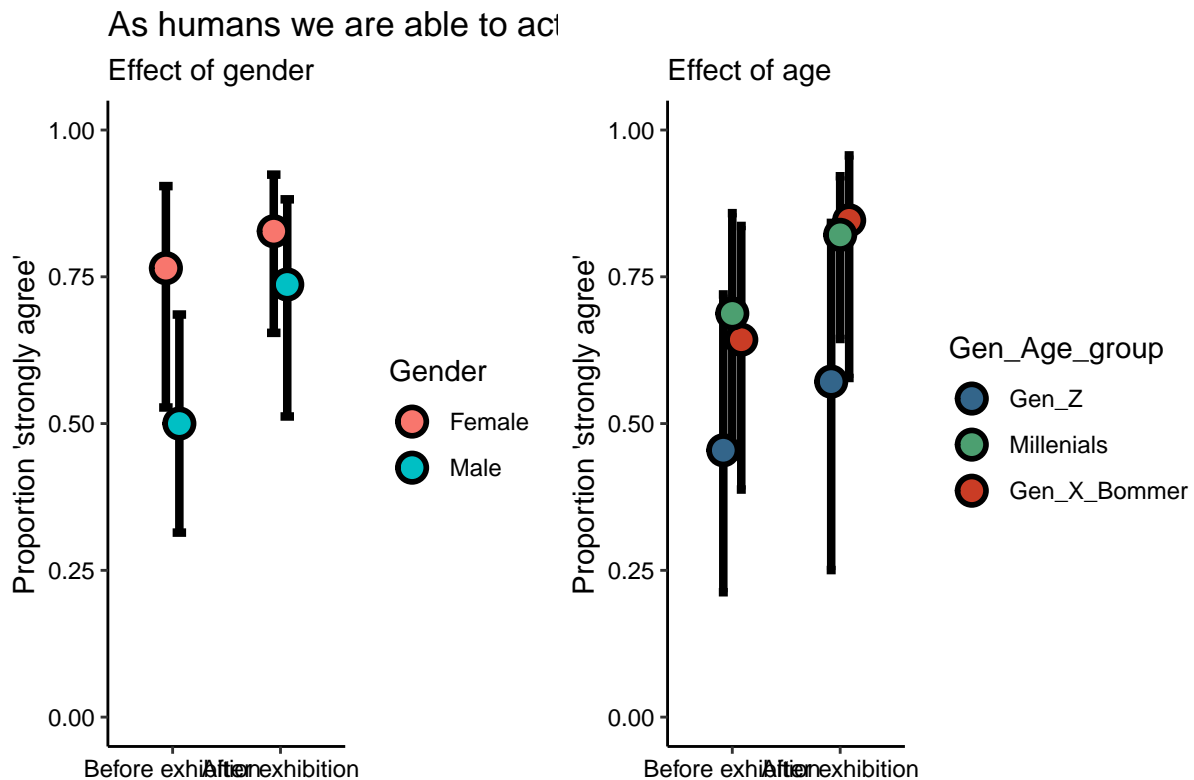

3. I have a personal responsibility to change my behaviour to avoid microplastic pollution

Step 1: Check whether data can be grouped for logistic regression.

```
ggplot() +
  geom_bar(data = clean_data, aes(x = Personal_responsibility_LK), colour = "black") +
  scale_x_continuous(name = "", limits = c(0.5, 7.5),
                     breaks = seq(1, 7, 1), labels = c("Strongly disagree", 2:6, "Strongly agree")) +
  scale_y_continuous(name = "Responses") +
  labs(title = "I have a personal responsibility to change my behaviour to avoid microplastic pollution
```

```
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

I have a personal responsibility to change my behaviour to avoid microplasti

There is a clear bias towards 'strongly agree' so we will analyse this data using a logistic regression (7 v. <7).

Step 2: Fit preliminary model

We include an interaction between gender and treatment to see if effect of exhibition has different effects on men and women. To start with we use age groups defined by 'generations'. We will check to see if this differs if we use age groups that are more balanced. We include Location to make sure there is no difference between Kiel and Berlin.

```
##Turn Likert scale into binary variable
Q3_data <- clean_data %>%
  dplyr::mutate(response_category = dplyr::case_when(Personal_responsibility_LK == 7 ~ 1,
                                                    Personal_responsibility_LK != 7 ~ 0))

##Create model
Q3_model <- glm(response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location,
            data = Q3_data, family = binomial(link = "logit"))

summary(Q3_model)
```

```
##
```

```
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Gen_Age_group + Location, family = binomial(link = "logit"),
##     data = Q3_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8700  -1.1539   0.6186   0.9564   1.8109
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                0.9314     0.8008   1.163   0.2448
## GenderMale                -1.6124     0.7108  -2.269   0.0233 *
## Gen_Age_groupGen_Z        -0.3806     0.8234  -0.462   0.6439
## Gen_Age_groupMillenials   -0.7430     0.7077  -1.050   0.2938
## LocationKIEL               0.6253     0.8959   0.698   0.4852
## GenderFemale:TreatmentTRUE 0.3565     0.6729   0.530   0.5963
## GenderMale:TreatmentTRUE   1.6129     0.6947   2.322   0.0203 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 122.02  on 88  degrees of freedom
## Residual deviance: 107.86  on 82  degrees of freedom
## AIC: 121.86
##
## Number of Fisher Scoring iterations: 4
```

Males are significantly less likely to strongly agree about personal responsibility; however, they are also significantly (positively) affected by the exhibition. Effect sizes of all other (non-significant) effects is much smaller.

Step 3: Check model reliability

Use DHARMa to check for issues of heteroskedasticity or outliers.

```
##Check model using DHARMa
simulationOutput <- simulateResiduals(Q3_model, seed = 123, n = 5000)
plot(simulationOutput)
```
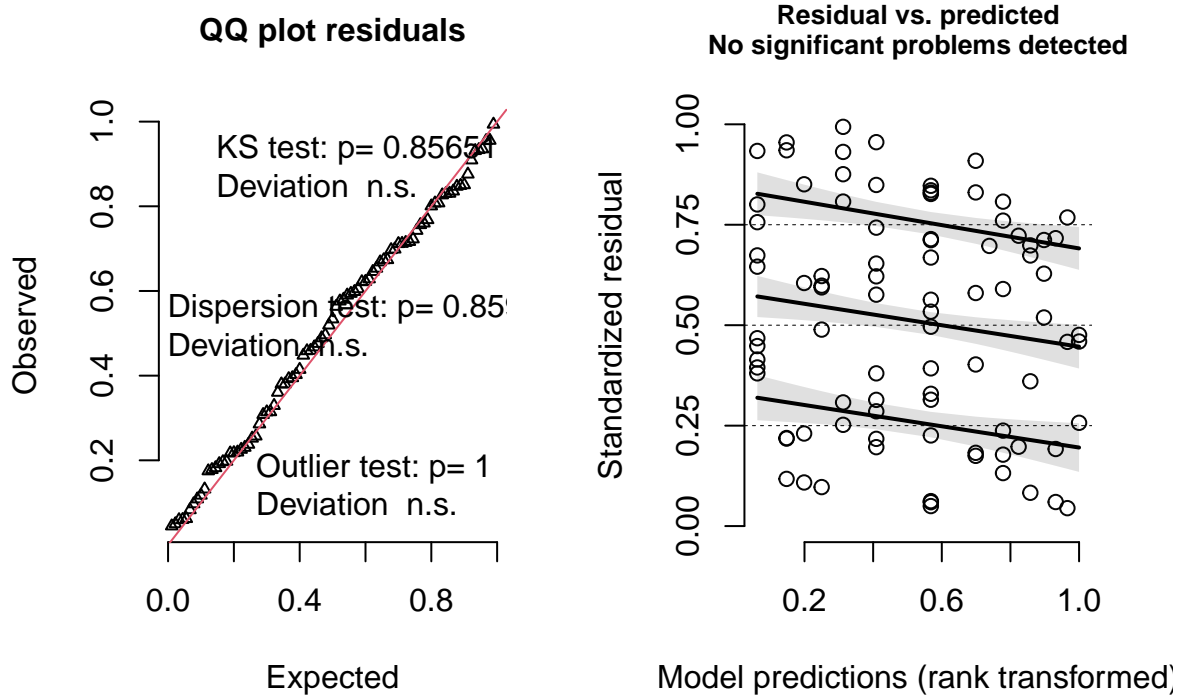
DHARMa residual diagnostics

QQ plot residuals

KS test: p= 0.85654
Deviation n.s.

Dispersion test: p= 0.85
Deviation n.s.

Outlier test: p= 1
Deviation n.s.

Observed
Expected

Residual vs. predicted
No significant problems detected

Standardized residual
Model predictions (rank transformed)

```r
#Check there are no unusual residual patterns for each predictor
plotResiduals(simulationOutput, form = as.factor(Q3_data$Gender))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
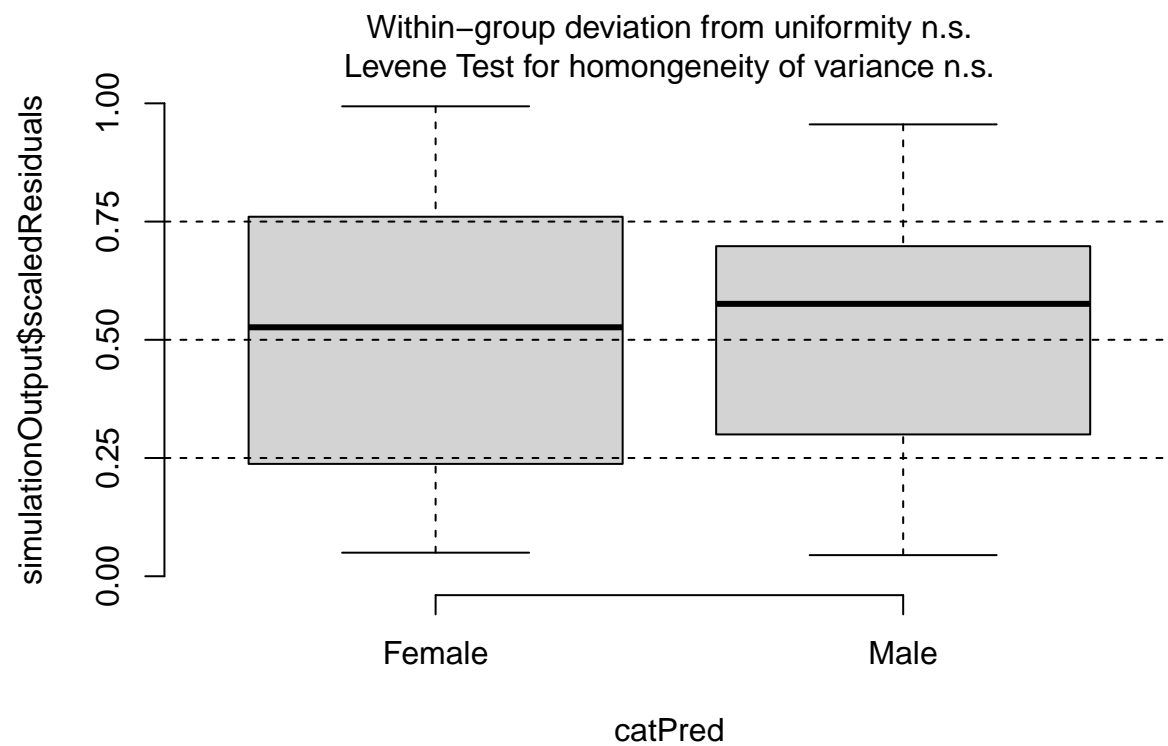
```
plotResiduals(simulationOutput, form = as.factor(Q3_data$Age_group))
```

Within-group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
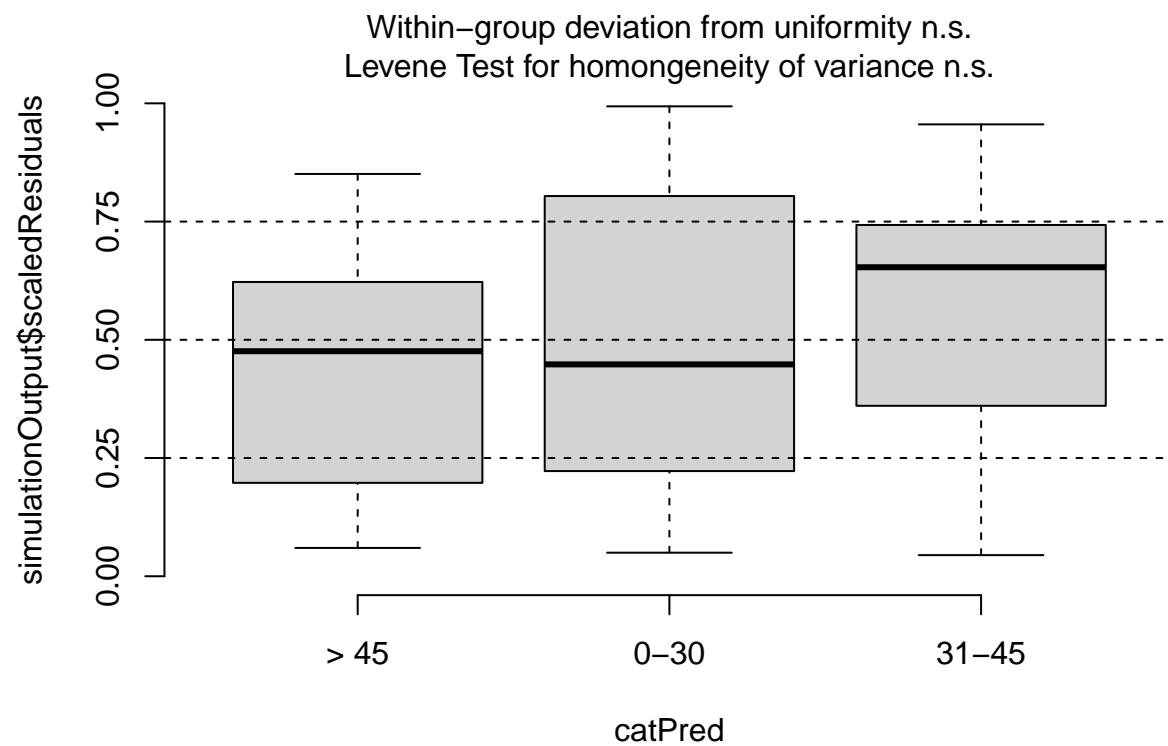
```
plotResiduals(simulationOutput, form = as.factor(Q3_data$Treatment))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
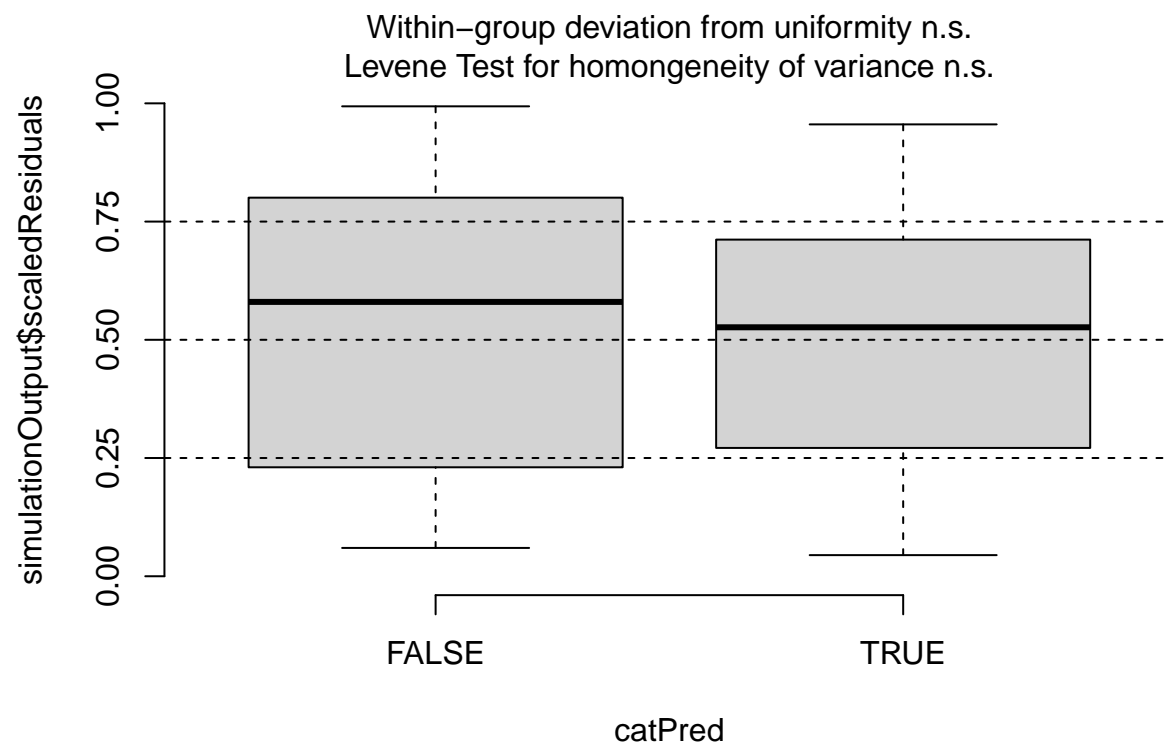
```
plotResiduals(simulationOutput, form = as.factor(Q3_data$Location))
```
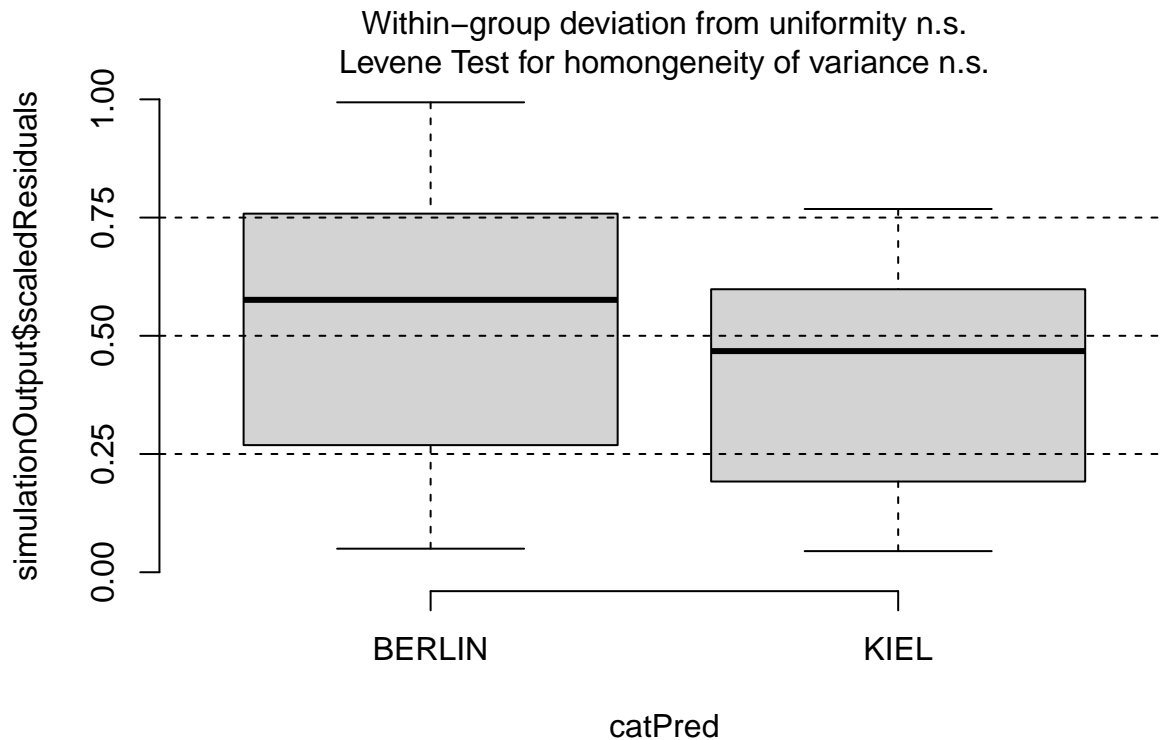
No apparent issues with model.

Step 3b: Check that fitting model with alternative age group gives the same results.

```
##Create model
Q3_model_age <- glm(response_category ~ Gender + Gender:Treatment + Age_group + Location,
                    data = Q3_data, family = binomial(link = "logit"))

summary(Q3_model_age)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Age_group + Location, family = binomial(link = "logit"),
##     data = Q3_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9275  -1.1851   0.6402   0.9880   1.8119
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.0435     0.8605   1.213   0.2252
## GenderMale                -1.6701     0.7163  -2.331   0.0197 *
```

```
## Age_group0-30                    -0.7998    0.7803  -1.025    0.3054
## Age_group31-45                   -0.3322    0.7989  -0.416    0.6775
## LocationKIEL                      0.6445    0.8744   0.737    0.4611
## GenderFemale:TreatmentTRUE        0.2196    0.6638   0.331    0.7407
## GenderMale:TreatmentTRUE          1.4630    0.7197   2.033    0.0421 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 122.02  on 88  degrees of freedom
## Residual deviance: 107.76  on 82  degrees of freedom
## AIC: 121.76
##
## Number of Fisher Scoring iterations: 4
```

Results are qualitatively the same. We can continue just using 'generational' cut-offs.

```
Q3_data %>%
  mutate(age_cutoff = cut(Age, breaks = c(seq(15, 45, 5), 85)),
         age_cutoff_numeric = as.numeric(age_cutoff)) %>%
  group_by(age_cutoff_numeric) %>%
  summarise(binom::binom.wilson(x = sum(response_category), n = n()),
            age_cutoff = first(age_cutoff),
            n = n()) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = age_cutoff_numeric, ymin = lower, ymax = upper), width = 0.25) +
      geom_point(aes(x = age_cutoff_numeric, y = mean), size = 3, shape = 21, fill = "grey45") +
      geom_text(aes(x = age_cutoff_numeric + 0.15, y = mean, label = n), size = 4) +
      scale_x_continuous(breaks = seq(1:max(.$age_cutoff_numeric)),
                         labels = .$age_cutoff) +
      scale_y_continuous(limits = c(0, 1)) +
      theme_classic()}
```

Could we even analyse this as a continous variable with quadratic effect?

```
##Create model
Q3_model_agequad <- glm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q3_data, family = binomial(link = "logit"))

summary(Q3_model_agequad)


##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     poly(Age, 2) + Location, family = binomial(link = "logit"),
##     data = Q3_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7321  -1.2755   0.5317   0.9696   1.7802
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               0.68262    0.56954   1.199   0.2307
## GenderMale               -1.65289    0.73068  -2.262   0.0237 *
## poly(Age, 2)1             5.50047    3.80361   1.446   0.1481
## poly(Age, 2)2             1.86209    2.77571   0.671   0.5023
## LocationKIEL              0.03941    0.99599   0.040   0.9684
## GenderFemale:TreatmentTRUE  0.21302    0.67231   0.317   0.7514
```
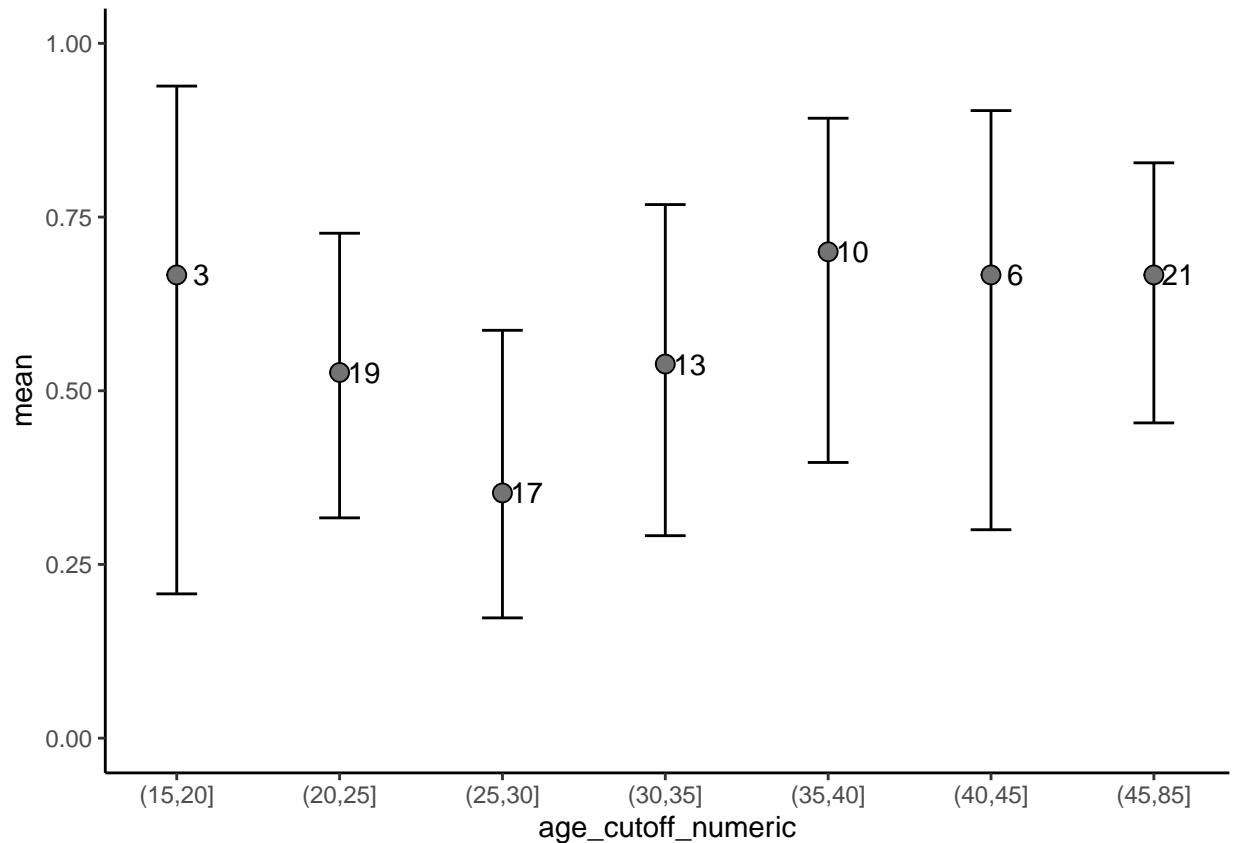
```
## GenderMale:TreatmentTRUE     1.56537     0.69579    2.250    0.0245 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 122.02  on 88  degrees of freedom
## Residual deviance: 106.53  on 82  degrees of freedom
## AIC: 120.53
##
## Number of Fisher Scoring iterations: 4
```

```
AIC(Q3_model, Q3_model_age, Q3_model_agequad)
```

```
##                  df      AIC
## Q3_model          7 121.8605
## Q3_model_age      7 121.7637
## Q3_model_agequad  7 120.5263
```

No major difference in AIC between models.

Step 4: Plot results

```
## Plot results (gender and treatment)
plot_data <- Q3_data %>%
  dplyr::group_by(Gender, Treatment) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(response_category), n = n()), .groups = "drop")

ggplot() +
  geom_errorbar(data = plot_data, aes(x = Treatment, ymin = lower, ymax = upper, group = Gender),
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data, aes(x = Treatment, y = mean, fill = Gender),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "I have a personal responsibility to change my behaviour to avoid microplastic pollution"
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

I have a personal responsibility to change my behaviour to avoid microplas
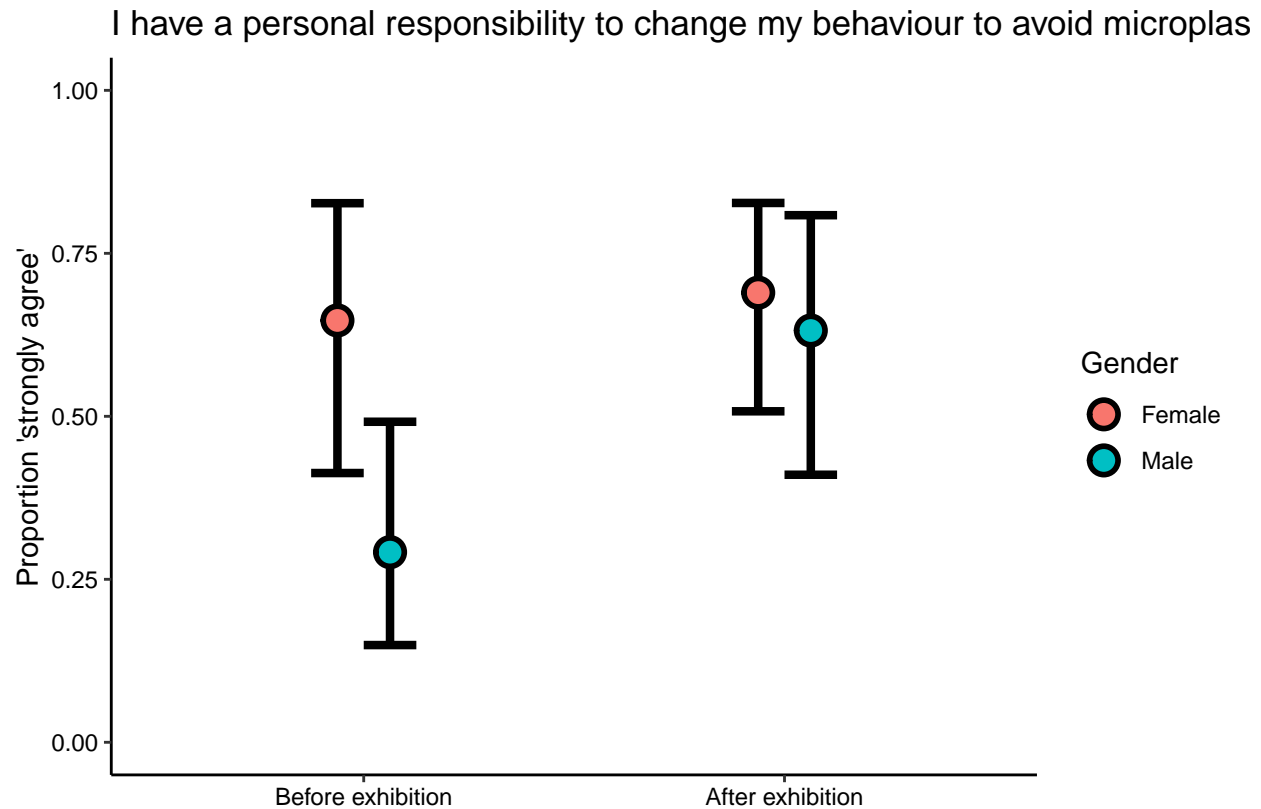
4. Politicians have a responsibility to implement legislation to avoid microplastic pollution

Step 1: Check whether data can be grouped for logistic regression.

```
ggplot() +
  geom_bar(data = clean_data, aes(x = Political_responsibility_LK), colour = "black") +
  scale_x_continuous(name = "", limits = c(0.5, 7.5),
                     breaks = seq(1, 7, 1), labels = c("Strongly disagree", 2:6, "Strongly agree")) +
  scale_y_continuous(name = "Responses") +
  labs(title = "Politicians have a responsibility to implement legislation to avoid microplastic polluti
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

Politicians have a responsibility to implement legislation to avoid microplasti

There is a clear bias towards 'strongly agree' so we will analyse this data using a logistic regression (7 v. <7).

Step 2: Fit preliminary model

We include an interaction between gender and treatment to see if effect of exhibition has different effects on men and women. To start with we use age groups defined by 'generations'. We will check to see if this differs if we use age groups that are more balanced. We include Location to make sure there is no difference between Kiel and Berlin.

```r
##Turn Likert scale into binary variable
Q4_data <- clean_data %>%
  dplyr::mutate(response_category = dplyr::case_when(Political_responsibility_LK == 7 ~ 1,
                                                    Political_responsibility_LK != 7 ~ 0))

##Create model
Q4_model <- glm(response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location,
                data = Q4_data, family = binomial(link = "logit"))

summary(Q4_model)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##      Gen_Age_group + Location, family = binomial(link = "logit"),
```

```
##      data = Q4_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0539   0.5124   0.5880   0.7229   0.9722
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  1.6883     0.9258   1.824   0.0682 .
## GenderMale                  -0.4797     0.8002  -0.599   0.5489
## Gen_Age_groupGen_Z          -0.2965     0.9126  -0.325   0.7453
## Gen_Age_groupMillenials     -0.1161     0.7725  -0.150   0.8805
## LocationKIEL                 0.0159     0.9487   0.017   0.9866
## GenderFemale:TreatmentTRUE   0.2757     0.8454   0.326   0.7444
## GenderMale:TreatmentTRUE    -0.5885     0.6827  -0.862   0.3887
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 94.841  on 88  degrees of freedom
## Residual deviance: 90.956  on 82  degrees of freedom
## AIC: 104.96
##
## Number of Fisher Scoring iterations: 4
```

No significant effects. All effect sizes are quite small.

Step 3: Check model reliability

Use DHARMa to check for issues of heteroskedasticity or outliers.

```
##Check model using DHARMa
simulationOutput <- simulateResiduals(Q4_model, seed = 123, n = 5000)
plot(simulationOutput)
```
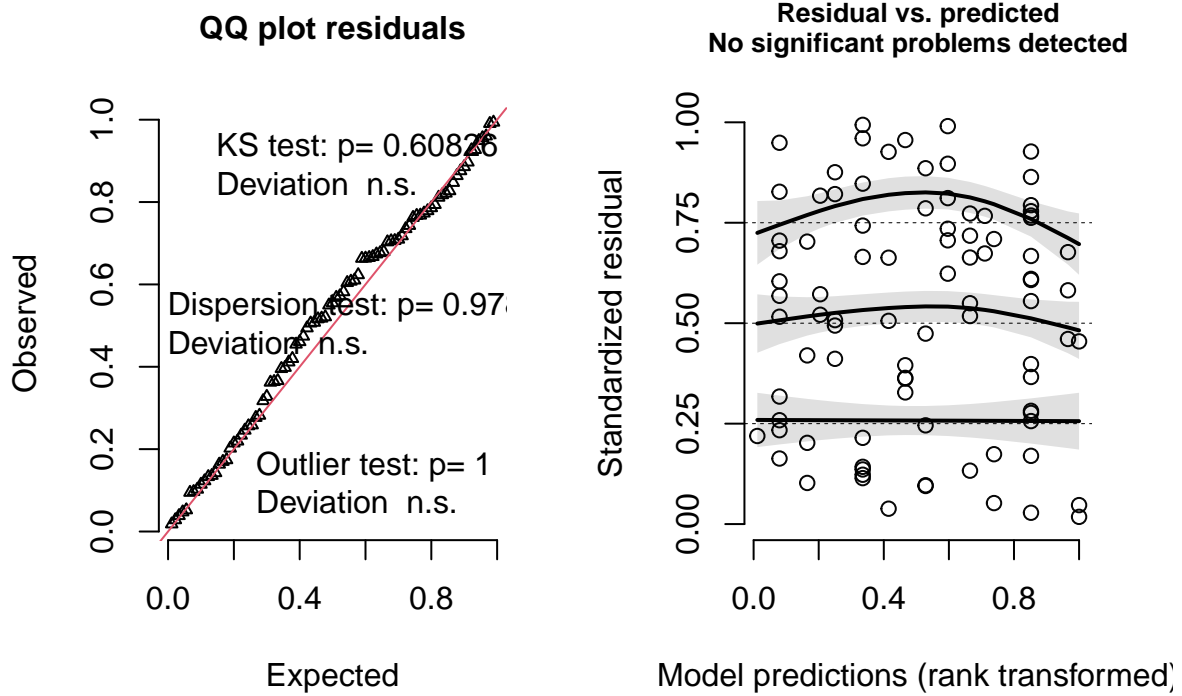
DHARMa residual diagnostics

QQ plot residuals

KS test: p= 0.60836
Deviation n.s.

Dispersion test: p= 0.97
Deviation n.s.

Outlier test: p= 1
Deviation n.s.

Residual vs. predicted
No significant problems detected


```
#Check there are no unusual residual patterns for each predictor
plotResiduals(simulationOutput, form = as.factor(Q4_data$Gender))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
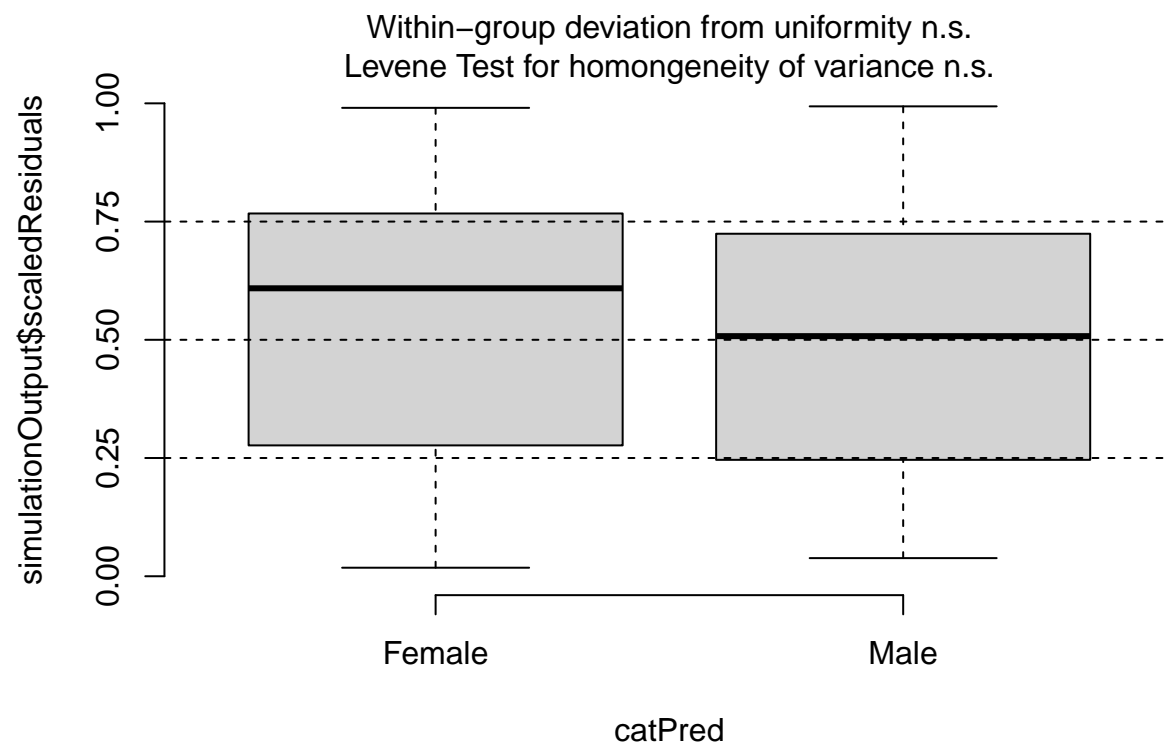
```
plotResiduals(simulationOutput, form = as.factor(Q4_data$Age_group))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
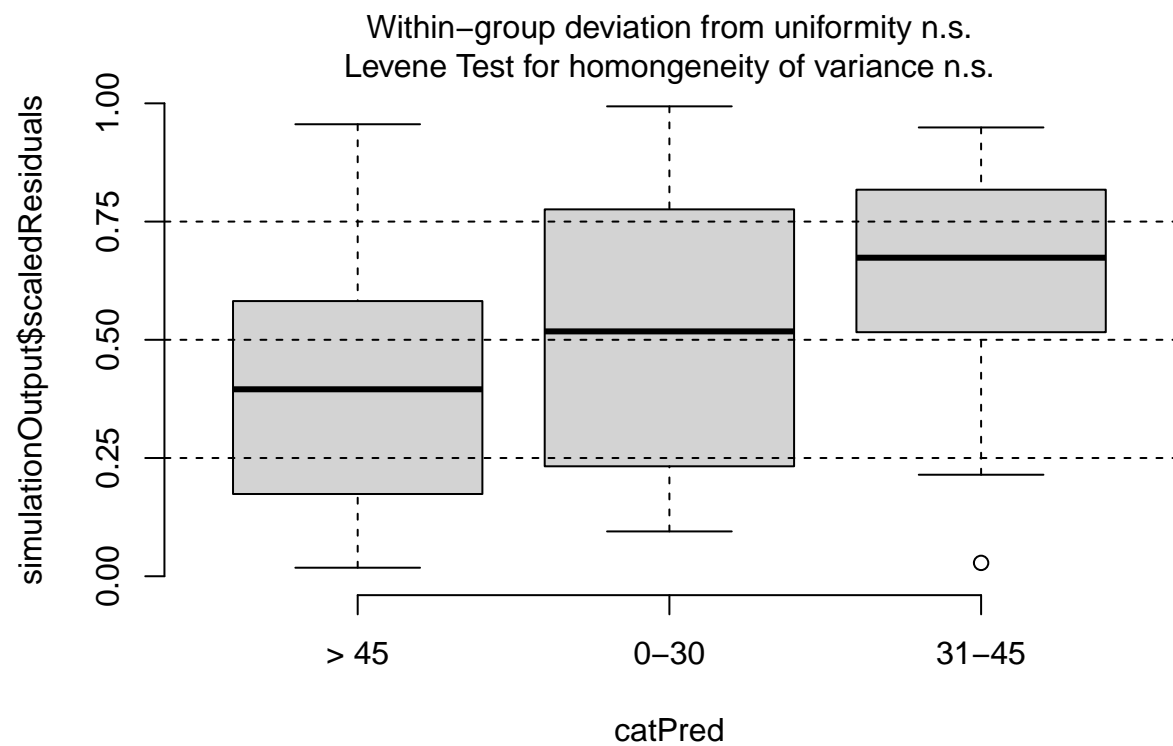
```
plotResiduals(simulationOutput, form = as.factor(Q4_data$Treatment))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
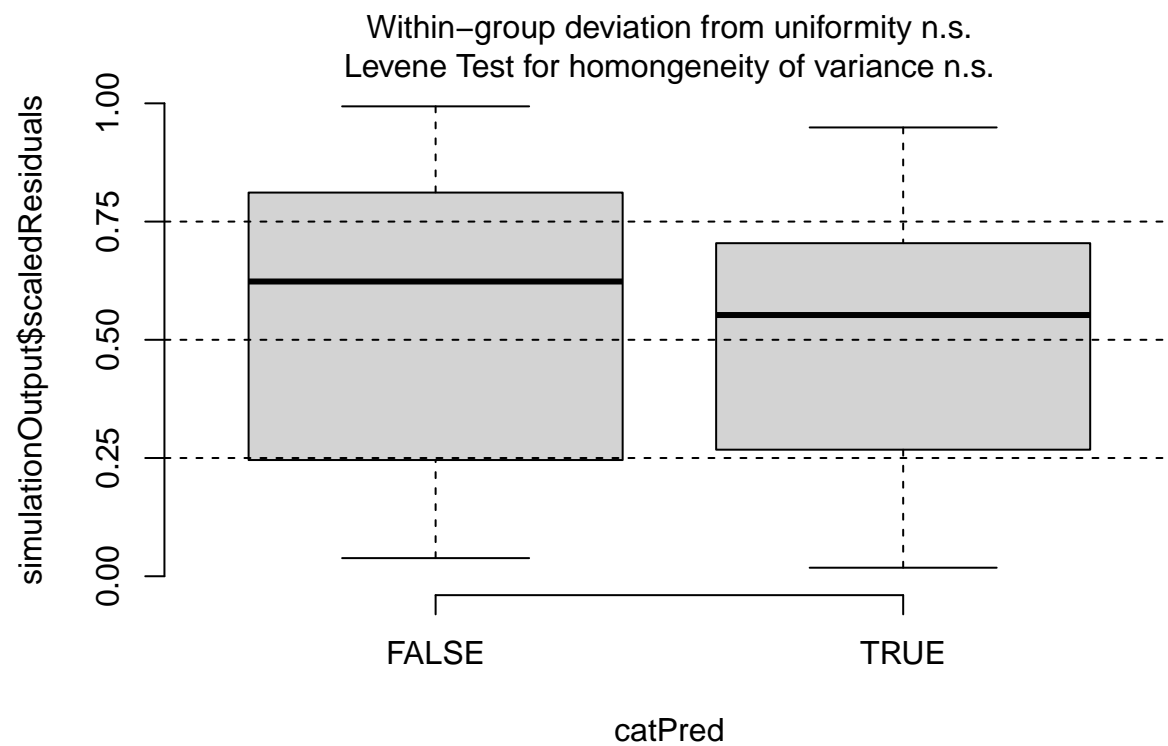
```
plotResiduals(simulationOutput, form = as.factor(Q4_data$Location))
```
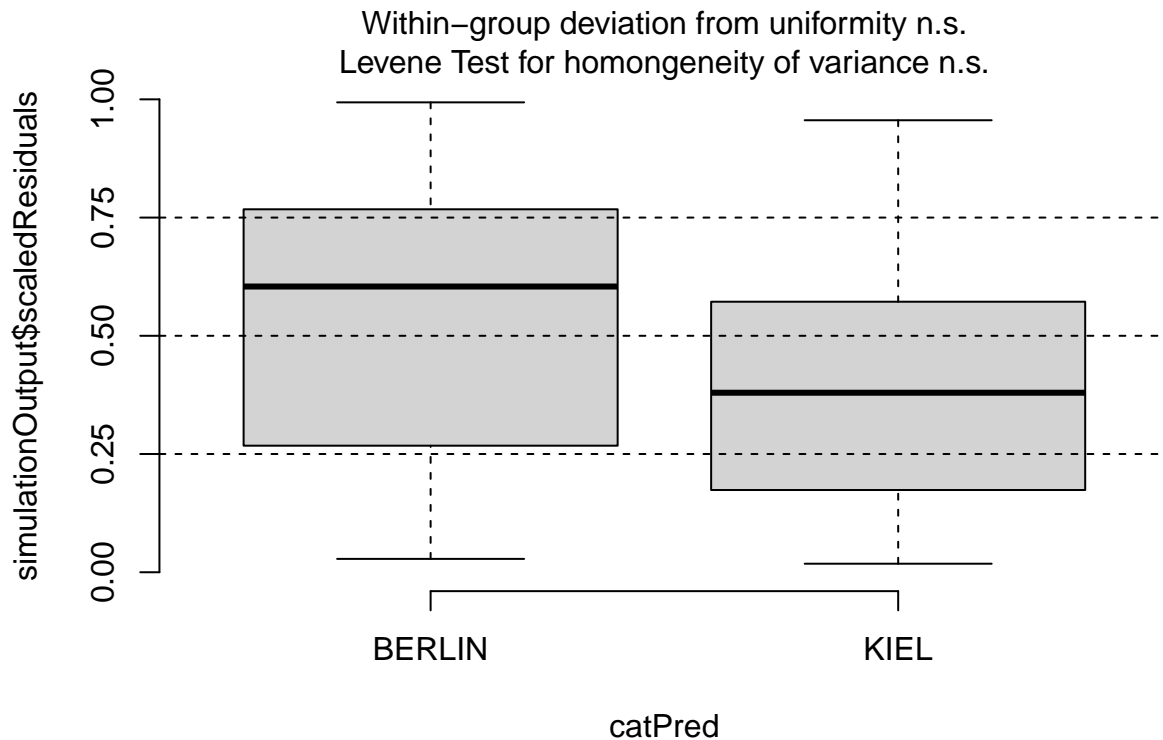
No apparent issues with model.

Step 3b: Check that fitting model with alternative age group gives the same results.

```
##Create model
Q4_model_age <- glm(response_category ~ Gender + Gender:Treatment + Age_group + Location,
                    data = Q4_data, family = binomial(link = "logit"))

summary(Q4_model_age)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Age_group + Location, family = binomial(link = "logit"),
##     data = Q4_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4340   0.3259   0.6163   0.7377   1.3000
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                0.9603     0.9332   1.029   0.3035
## GenderMale                -0.3640     0.8065  -0.451   0.6518
```

```
## Age_group0-30                0.3427    0.7921   0.433   0.6653
## Age_group31-45               1.6871    0.9165   1.841   0.0656 .
## LocationKIEL                 0.5662    0.9131   0.620   0.5352
## GenderFemale:TreatmentTRUE   0.2616    0.8494   0.308   0.7581
## GenderMale:TreatmentTRUE    -1.2226    0.7848  -1.558   0.1193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 94.841  on 88  degrees of freedom
## Residual deviance: 86.265  on 82  degrees of freedom
## AIC: 100.27
##
## Number of Fisher Scoring iterations: 4
```

Results are qualitatively the same. Some suggestion of effect in 31-45yo group, but not significant. We will continue just using 'generational' cut-offs.

```
Q4_data %>%
  mutate(age_cutoff = cut(Age, breaks = c(seq(15, 45, 5), 85)),
         age_cutoff_numeric = as.numeric(age_cutoff)) %>%
  group_by(age_cutoff_numeric) %>%
  summarise(binom::binom.wilson(x = sum(response_category), n = n()),
            age_cutoff = first(age_cutoff),
            n = n()) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = age_cutoff_numeric, ymin = lower, ymax = upper), width = 0.25) +
      geom_point(aes(x = age_cutoff_numeric, y = mean), size = 3, shape = 21, fill = "grey45") +
      geom_text(aes(x = age_cutoff_numeric + 0.15, y = mean, label = n), size = 4) +
      scale_x_continuous(breaks = seq(1:max(.$age_cutoff_numeric)),
                         labels = .$age_cutoff) +
      scale_y_continuous(limits = c(0, 1)) +
      theme_classic()}
```

Could we even analyse this as a continous variable with quadratic effect?

```
##Create model
Q4_model_agequad <- glm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q4_data, family = binomial(link = "logit"))

summary(Q4_model_agequad)
```
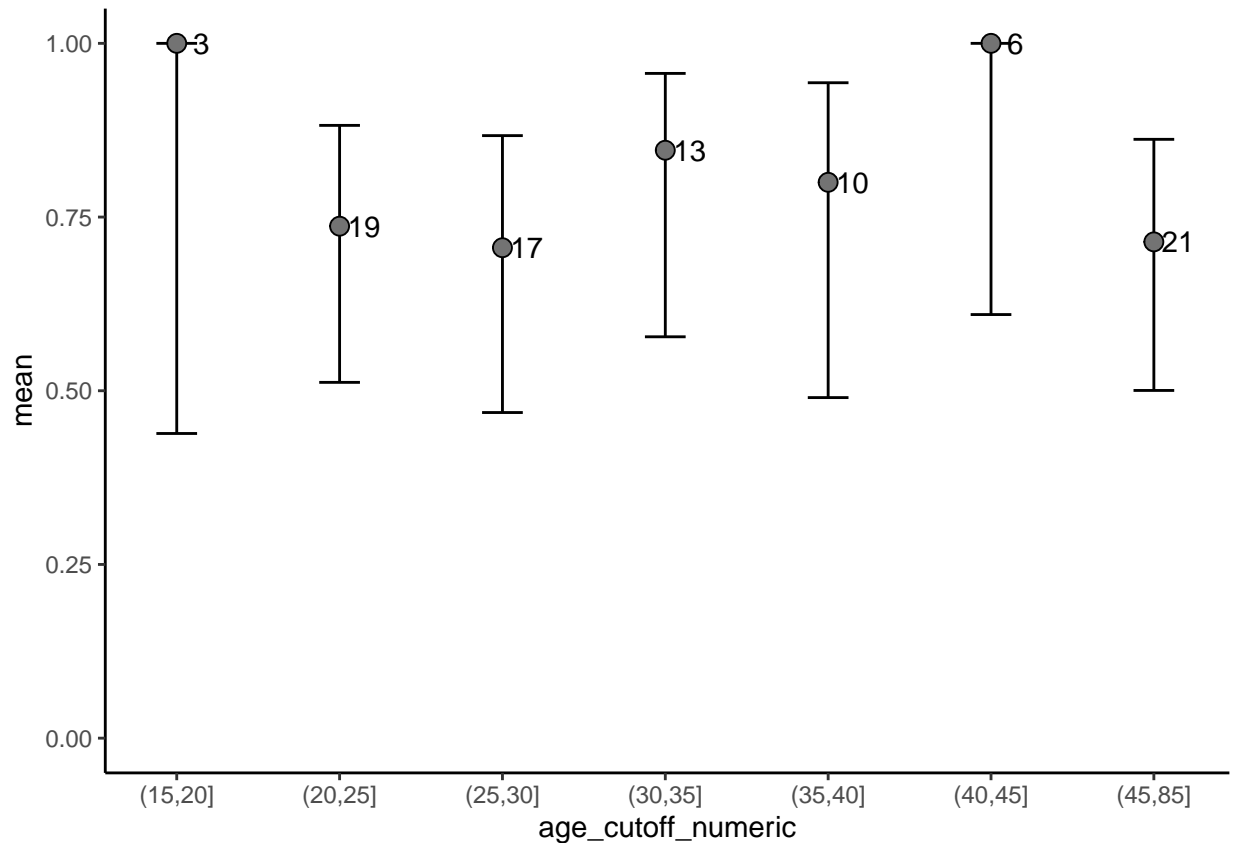
```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     poly(Age, 2) + Location, family = binomial(link = "logit"),
##     data = Q4_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1182   0.4992   0.5864   0.7138   1.0045
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.6132     0.6912   2.334   0.0196 *
## GenderMale                -0.5504     0.8139  -0.676   0.4989
## poly(Age, 2)1              0.1059     3.4153   0.031   0.9753
## poly(Age, 2)2             -1.6669     2.7083  -0.615   0.5382
## LocationKIEL               0.2022     1.0171   0.199   0.8424
## GenderFemale:TreatmentTRUE 0.2215    0.8505   0.260   0.7945
```

```
## GenderMale:TreatmentTRUE    -0.6562    0.6976  -0.941    0.3469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 94.841  on 88  degrees of freedom
## Residual deviance: 90.685  on 82  degrees of freedom
## AIC: 104.69
##
## Number of Fisher Scoring iterations: 4
```

```
AIC(Q4_model, Q4_model_age, Q4_model_agequad)
```

```
##                  df      AIC
## Q4_model          7 104.9562
## Q4_model_age      7 100.2654
## Q4_model_agequad  7 104.6855
```

In this case, using age groups (not generations) is noticably better using AIC.

Step 4: Plot results

```
## Plot results (gender and treatment)
plot_data <- Q4_data %>%
  dplyr::group_by(Gender, Treatment) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(response_category), n = n()), .groups = "drop")

ggplot() +
  geom_errorbar(data = plot_data, aes(x = Treatment, ymin = lower, ymax = upper, group = Gender),
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data, aes(x = Treatment, y = mean, fill = Gender),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "Politicians have a responsibility to implement legislation to avoid microplastic polluti
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

Politicians have a responsibility to implement legislation to avoid microplast
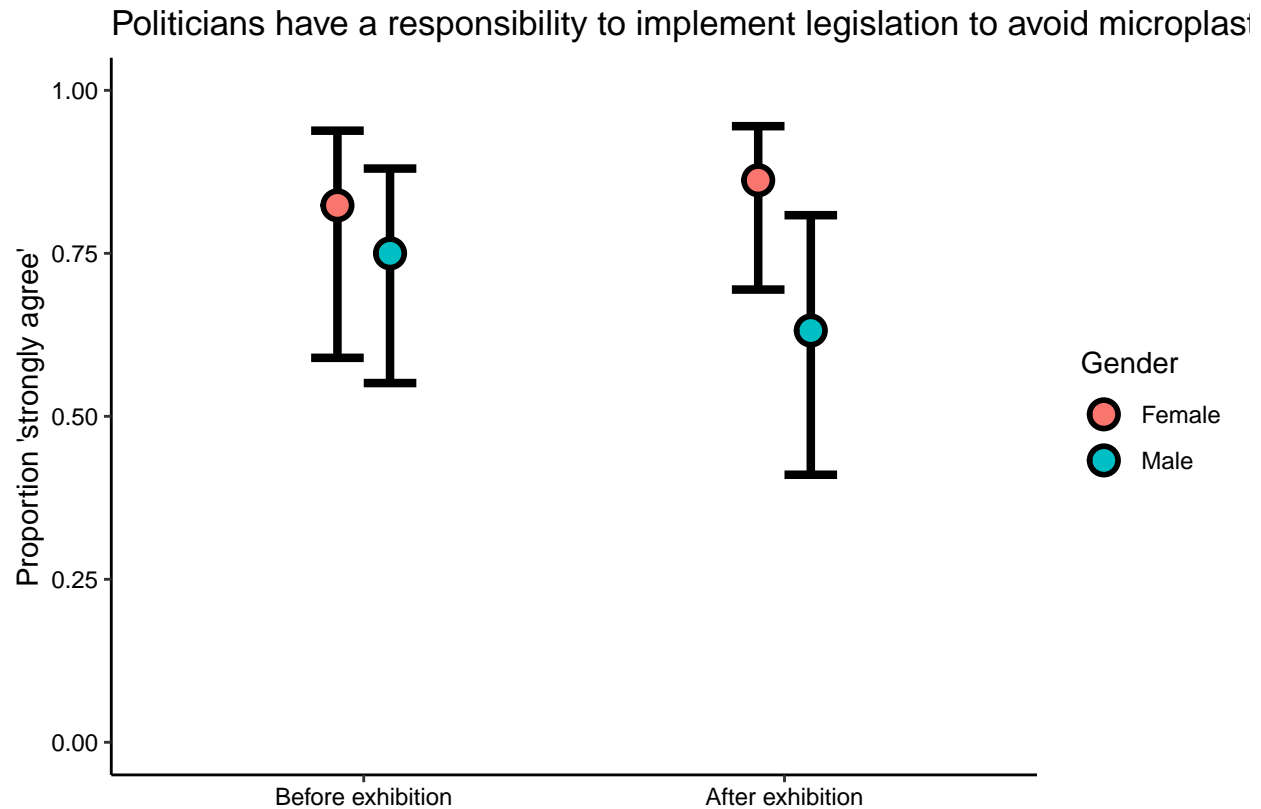
5. Climate change is the most important problem facing our society

Step 1: Check whether data can be grouped for logistic regression.

```
ggplot() +
  geom_bar(data = clean_data, aes(x = Climate_importance_LK), colour = "black") +
  scale_x_continuous(name = "", limits = c(0.5, 7.5),
                     breaks = seq(1, 7, 1), labels = c("Strongly disagree", 2:6, "Strongly agree")) +
  scale_y_continuous(name = "Responses") +
  labs(title = "Climate change is the most important problem facing our society") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

Climate change is the most important problem facing our society

There is a clear bias towards 'strongly agree' so we will analyse this data using a logistic regression (7 v. <7).

Step 2: Fit preliminary model

We include an interaction between gender and treatment to see if effect of exhibition has different effects on men and women. To start with we use age groups defined by 'generations'. We will check to see if this differs if we use age groups that are more balanced. We include Location to make sure there is no difference between Kiel and Berlin.

```
##Turn Likert scale into binary variable
Q5_data <- clean_data %>%
  dplyr::mutate(response_category = dplyr::case_when(Climate_importance_LK == 7 ~ 1,
                                                     Climate_importance_LK != 7 ~ 0))


##Create model
Q5_model <- glm(response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location,
                data = Q5_data, family = binomial(link = "logit"))


summary(Q5_model)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Gen_Age_group + Location, family = binomial(link = "logit"),
```

```
##      data = Q5_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7063  -1.1872  -0.4227   0.9995   1.9325
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -1.2033     0.9085  -1.324   0.1853
## GenderMale                -1.1671     0.6989  -1.670   0.0949 .
## Gen_Age_groupGen_Z         1.3893     0.9295   1.495   0.1350
## Gen_Age_groupMillenials    1.8664     0.8388   2.225   0.0261 *
## LocationKIEL               2.3935     0.9721   2.462   0.0138 *
## GenderFemale:TreatmentTRUE -0.2290    0.6594  -0.347   0.7283
## GenderMale:TreatmentTRUE    0.6710    0.6768   0.991   0.3215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 123.28  on 88  degrees of freedom
## Residual deviance: 111.13  on 82  degrees of freedom
## AIC: 125.13
##
## Number of Fisher Scoring iterations: 4
```

Millenials are significantly more likely to care about climate change than GenX+. GenZ also has similar effect size, although not significant. People in Kiel were much more likely to care about climate change than in Berlin. Males still appear to be less concerned than women overall. No effect of treatment.

Step 3: Check model reliability

Use DHARMa to check for issues of heteroskedasticity or outliers.

```
##Check model using DHARMa
simulationOutput <- simulateResiduals(Q5_model, seed = 123, n = 5000)
plot(simulationOutput)
```
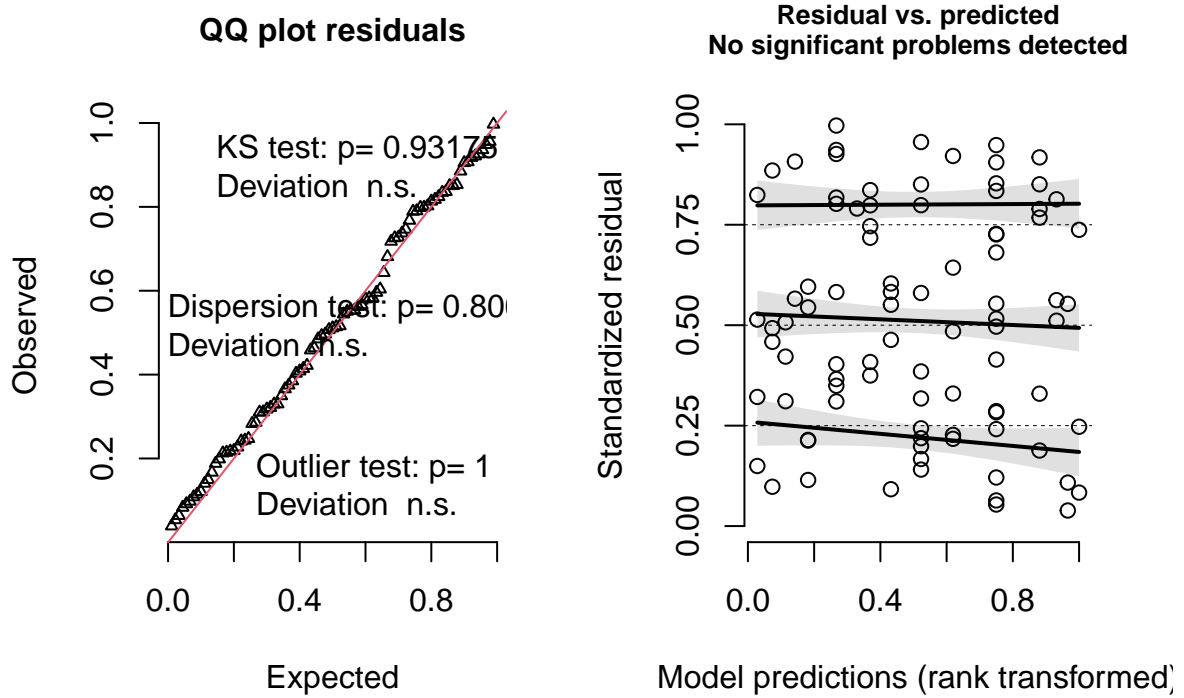
DHARMa residual diagnostics

QQ plot residuals

KS test: p= 0.9317
Deviation n.s.

Dispersion test: p= 0.80
Deviation n.s.

Outlier test: p= 1
Deviation n.s.

Observed

Expected

Residual vs. predicted
No significant problems detected

Standardized residual

Model predictions (rank transformed)

```
#Check there are no unusual residual patterns for each predictor
plotResiduals(simulationOutput, form = as.factor(Q5_data$Gender))
```

Within-group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
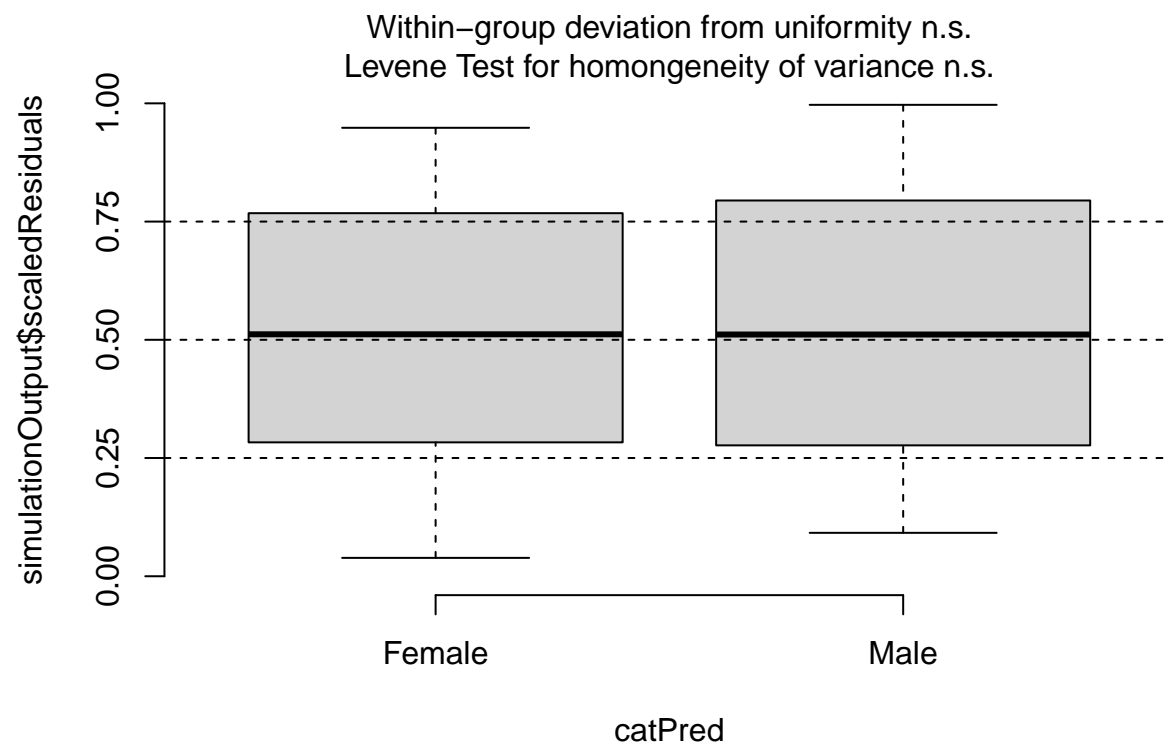
```
plotResiduals(simulationOutput, form = as.factor(Q5_data$Age_group))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
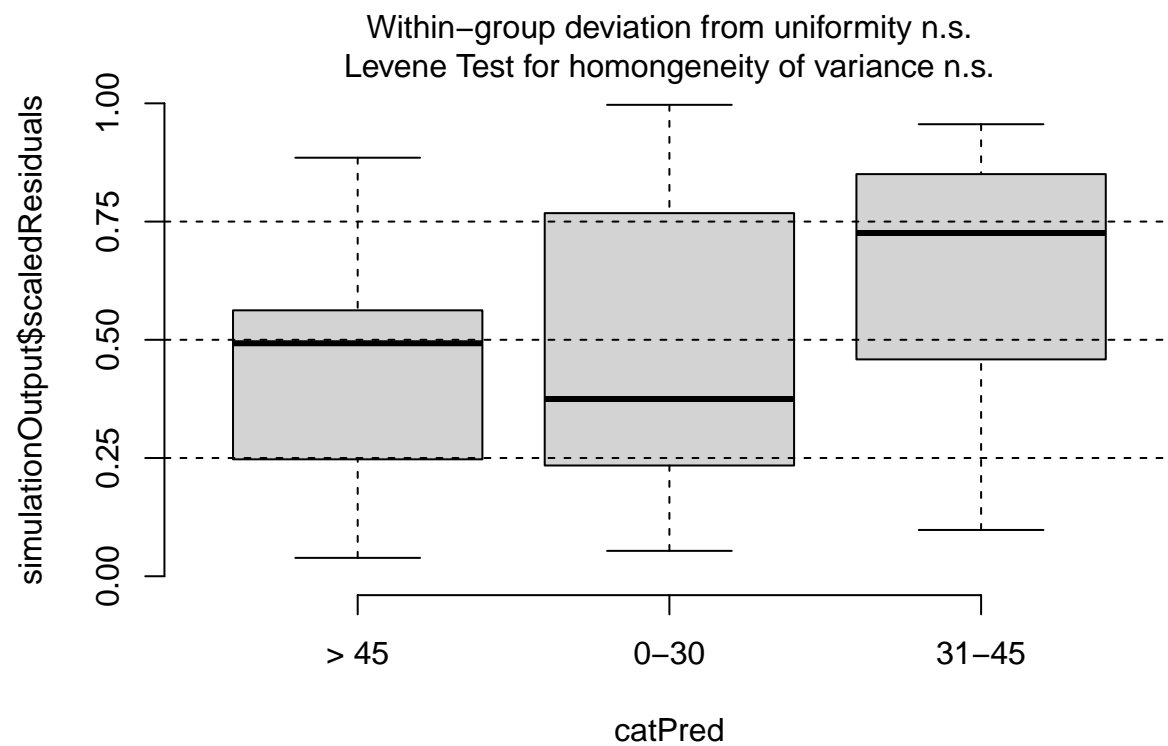
```
plotResiduals(simulationOutput, form = as.factor(Q5_data$Treatment))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
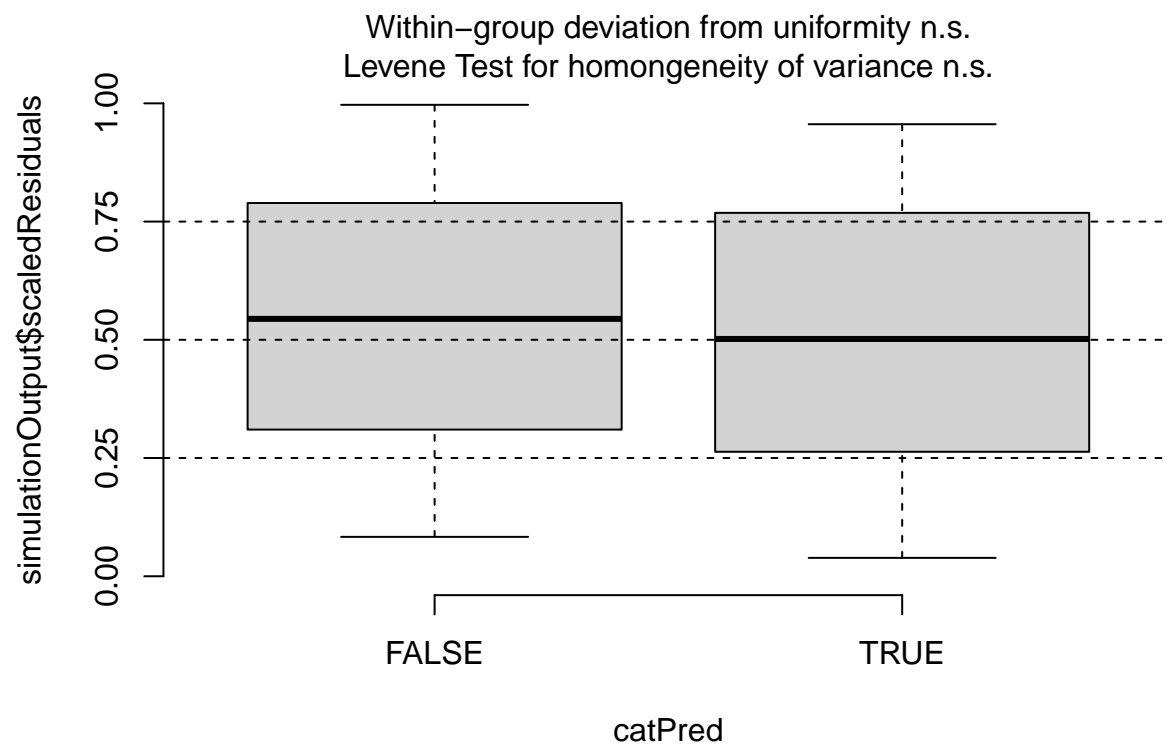
```
plotResiduals(simulationOutput, form = as.factor(Q5_data$Location))
```
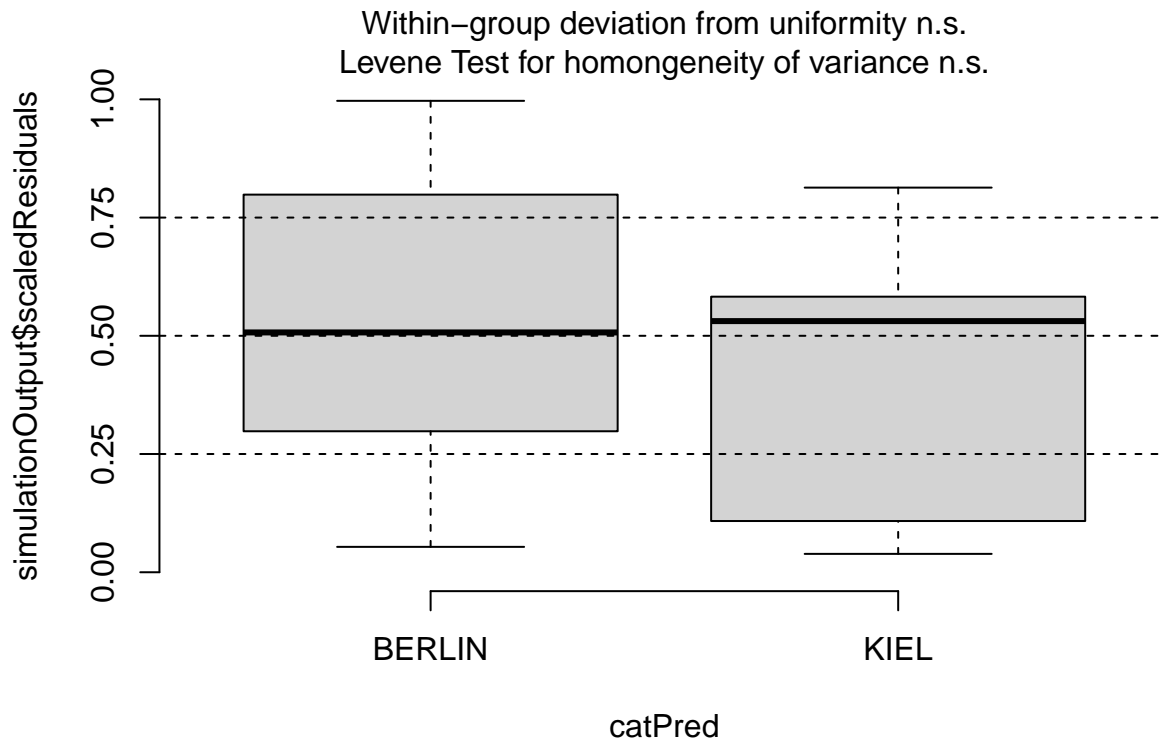
Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.

No apparent issues with model.

Step 3b: Check that fitting model with alternative age group gives the same results.

```
##Create model
Q5_model_age <- glm(response_category ~ Gender + Gender:Treatment + Age_group + Location,
                    data = Q5_data, family = binomial(link = "logit"))

summary(Q5_model_age)
```

```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     Age_group + Location, family = binomial(link = "logit"),
##     data = Q5_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7924  -1.1159  -0.3598   1.0792   2.3930
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.65725    1.17033  -1.416  0.15676
## GenderMale          -1.04783    0.70356  -1.489  0.13640
```

```
## Age_group0-30                    1.64436     1.11739    1.472  0.14113
## Age_group31-45                   3.03971     1.15187    2.639  0.00832 **
## LocationKIEL                     2.62053     1.19298    2.197  0.02805 *
## GenderFemale:TreatmentTRUE  -0.13358     0.66709   -0.200  0.84129
## GenderMale:TreatmentTRUE    -0.09926     0.75216   -0.132  0.89501
## ---
## Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 123.28  on 88   degrees of freedom
## Residual deviance: 104.60  on 82   degrees of freedom
## AIC: 118.6
##
## Number of Fisher Scoring iterations: 5
```

Results are qualitatively the same. We can continue just using 'generational' cut-offs.

```
Q5_data %>%
  mutate(age_cutoff = cut(Age, breaks = c(seq(15, 45, 5), 85)),
         age_cutoff_numeric = as.numeric(age_cutoff)) %>%
  group_by(age_cutoff_numeric) %>%
  summarise(binom::binom.wilson(x = sum(response_category), n = n()),
            age_cutoff = first(age_cutoff),
            n = n()) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = age_cutoff_numeric, ymin = lower, ymax = upper), width = 0.25) +
      geom_point(aes(x = age_cutoff_numeric, y = mean), size = 3, shape = 21, fill = "grey45") +
      geom_text(aes(x = age_cutoff_numeric + 0.15, y = mean, label = n), size = 4) +
      scale_x_continuous(breaks = seq(1:max(.$age_cutoff_numeric)),
                         labels = .$age_cutoff) +
      scale_y_continuous(limits = c(0, 1)) +
      theme_classic()}
```

Could we even analyse this as a continous variable with quadratic effect?

```
##Create model
Q5_model_agequad <- glm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q5_data, family = binomial(link = "logit"))

summary(Q5_model_agequad)
```
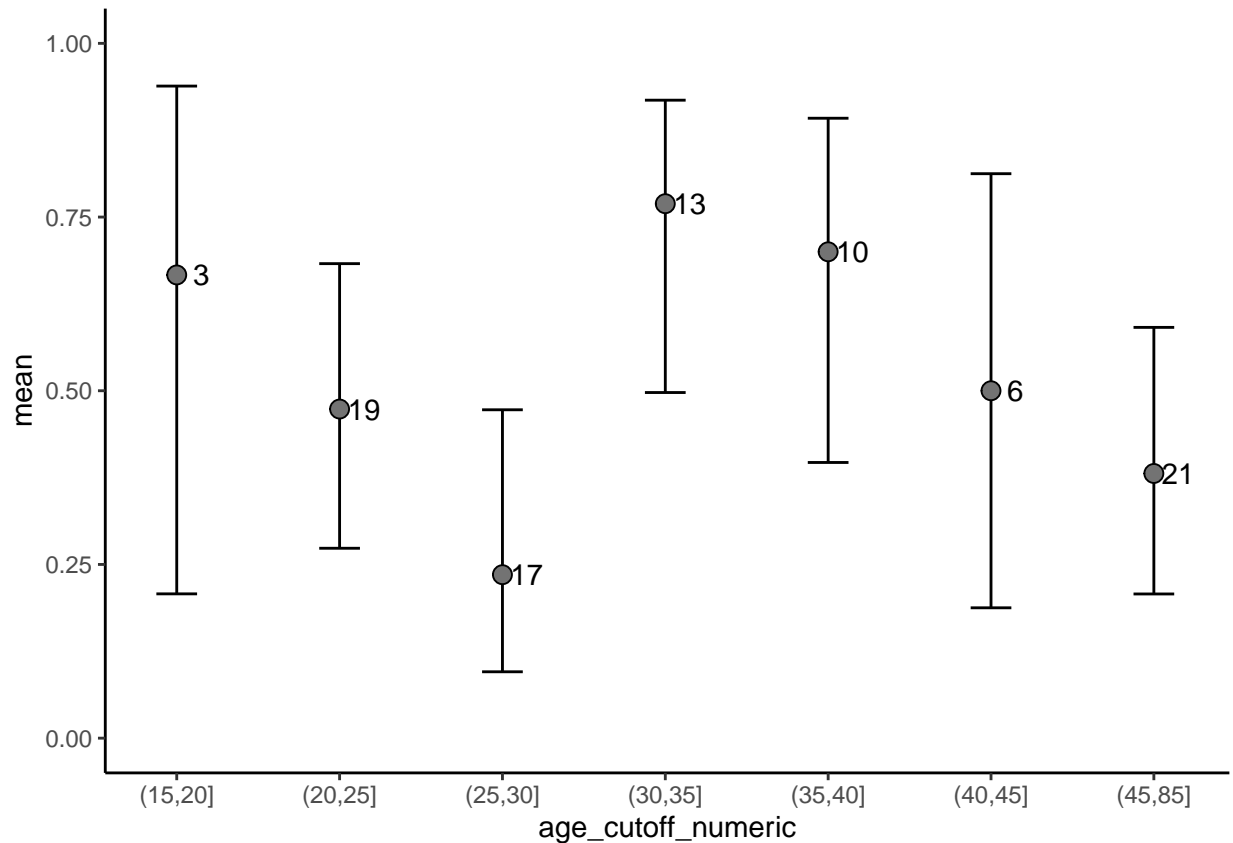
```
##
## Call:
## glm(formula = response_category ~ Gender + Gender:Treatment +
##     poly(Age, 2) + Location, family = binomial(link = "logit"),
##     data = Q5_data)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q     Max
## -1.9855  -1.0658  -0.5578   1.0835   1.5788
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               0.08304    0.54048    0.154   0.8779
## GenderMale               -1.23194    0.69117   -1.782   0.0747 .
## poly(Age, 2)1            -5.23383    3.30818   -1.582   0.1136
## poly(Age, 2)2            -2.61258    2.41318   -1.083   0.2790
## LocationKIEL              2.16472    1.02060    2.121   0.0339 *
## GenderFemale:TreatmentTRUE -0.11866  0.64373   -0.184   0.8537
```

```
## GenderMale:TreatmentTRUE     0.63894      0.67840    0.942    0.3463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 123.28  on 88  degrees of freedom
## Residual deviance: 113.93  on 82  degrees of freedom
## AIC: 127.93
##
## Number of Fisher Scoring iterations: 4
```

```
AIC(Q5_model, Q5_model_age, Q5_model_agequad)
```

```
##                   df       AIC
## Q5_model           7 125.1284
## Q5_model_age       7 118.5957
## Q5_model_agequad   7 127.9307
```

Again, using the balanced age choices is much better using AIC.

Step 4: Plot results

```
## Plot results (gender and treatment)
plot_data <- Q5_data %>%
  dplyr::group_by(Gender, Treatment) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(response_category), n = n()), .groups = "drop")

ggplot() +
  geom_errorbar(data = plot_data, aes(x = Treatment, ymin = lower, ymax = upper, group = Gender),
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data, aes(x = Treatment, y = mean, fill = Gender),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "Climate change is the most important problem facing our society") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```

Climate change is the most important problem facing our society
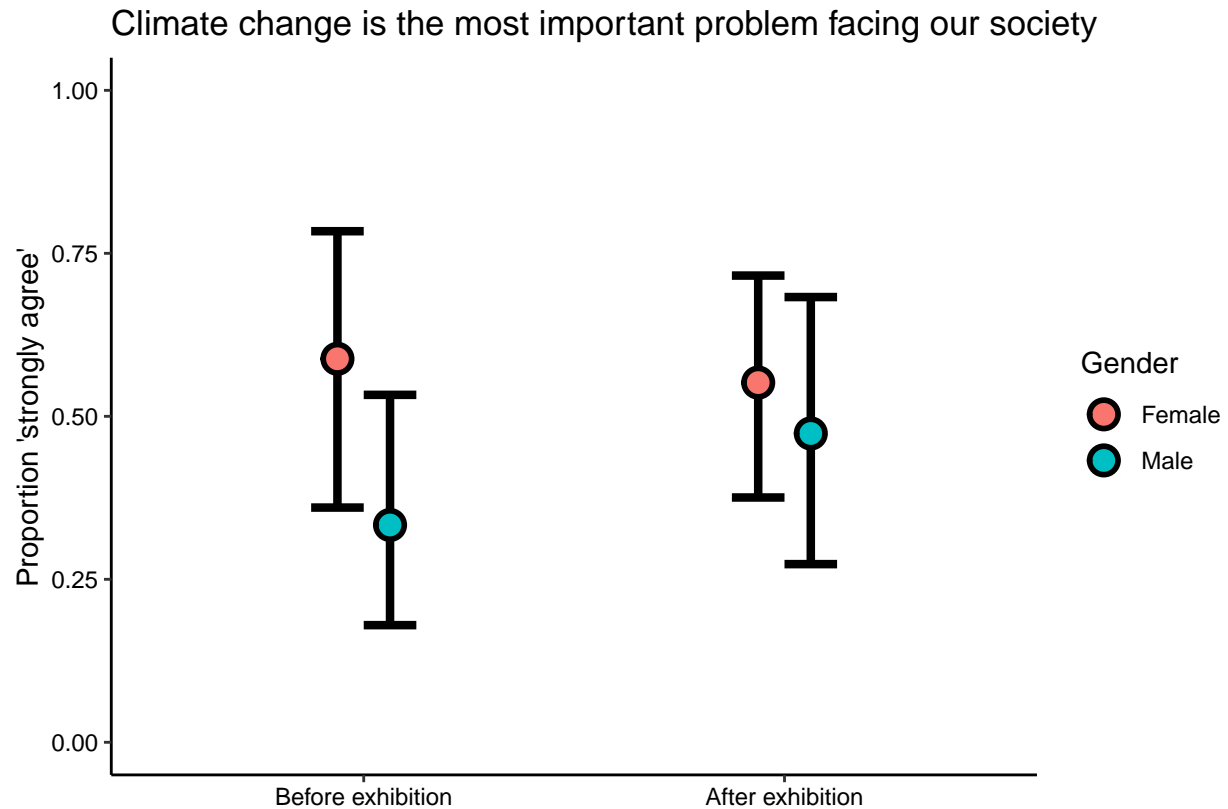
6. I would be willing to pay 50% more for products that do not contain microplastics

Step 1: Check whether data can be grouped for logistic regression.

```
ggplot() +
  geom_bar(data = clean_data, aes(x = Willingness_price_LK), colour = "black") +
  scale_x_continuous(name = "", limits = c(0.5, 7.5),
                     breaks = seq(1, 7, 1), labels = c("Strongly disagree", 2:6, "Strongly agree")) +
  scale_y_continuous(name = "Responses") +
  labs(title = "I would be willing to pay 50% more for products that do not contain microplastics") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```
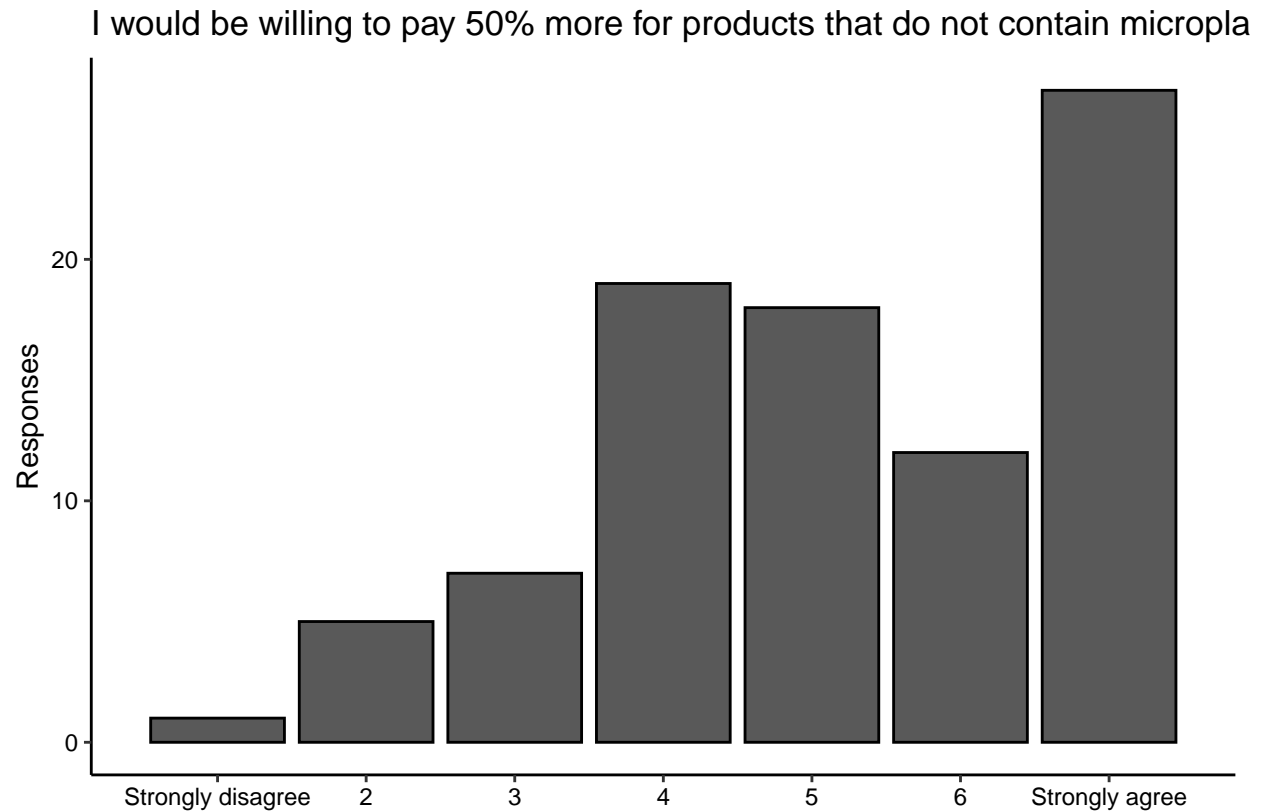
I would be willing to pay 50% more for products that do not contain micropla

Responses are more evenly spread than previous questions. Instead we will need to analyse using ordinal logistic regression (1-2; 3-5; 6-7).

Step 2: Fit preliminary model

We include an interaction between gender and treatment to see if effect of exhibition has different effects on men and women. To start with we use age groups defined by 'generations'. We will check to see if this differs if we use age groups that are more balanced. We include Location to make sure there is no difference between Kiel and Berlin.

```r
Q6_data <- clean_data %>%
  dplyr::mutate(response_category = dplyr::case_when(Willingness_price_LK < 3 ~ "Disagree",
                                                     Willingness_price_LK >= 3 & Willingness_price_LK <
                                                     Willingness_price_LK >= 6 ~ "Agree")) %>%
  #Make the new category ordered
  dplyr::mutate(response_category = ordered(response_category, levels = c("Disagree", "Neutral", "Agree
                across(.cols = c(Gen_Age_group, Location, Gender), .fns = as.factor))

##Create model
Q6_model <- clm(response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location,
                data = Q6_data, link = "logit")
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
summary(Q6_model)
```

```
## formula:
## response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location
## data:    Q6_data
##
##  link  threshold nobs logLik AIC    niter max.grad cond.H
##  logit flexible  89   -76.72 169.44 6(0)  2.34e-13 8.0e+01
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## GenderMale                -0.9012     0.6475  -1.392    0.164
## Gen_Age_groupGen_Z         0.3563     0.7548   0.472    0.637
## Gen_Age_groupMillenials    0.3156     0.6386   0.494    0.621
## LocationKIEL               0.6331     0.7783   0.813    0.416
## GenderFemale:TreatmentTRUE 0.2204     0.6197   0.356    0.722
## GenderMale:TreatmentTRUE   0.7869     0.6211   1.267    0.205
##
## Threshold coefficients:
##                 Estimate Std. Error z value
## Disagree|Neutral -2.6075     0.8379  -3.112
## Neutral|Agree     0.3972     0.7614   0.522
```

Interpretation: The 'threshold coefficients' signify the probability (on logit scale) that a response is less than or equal to a given value. So, the value given for Disagree|Neutral is the probability (on logit scale) that somebody would answer 'Disagree' (it is the lowest category so it is only one). The value given for Neutral|Agree is the probability (on logit scale) that somebody would answer 'Disagree' or 'Neutral'.

Note that these are the *intercept* values. The exact probabilities for a given respondent will vary based on the other variables in the model (which can be interpreted the same as a regular logistic regression). Remember, the assumption of proportional odds means we are assuming the effect of each variable is the same regardless of if we're looking at Disagree|Neutral or Neutral|Agree. We check this assumption below.

So, let's consider a male Gen Z respondent from Berlin that hasn't seen the exhibition.

```
thresholds  <- coefficients(Q6_model)[1:2]
gender_coef <- coefficients(Q6_model)[3]
genz_coef   <- coefficients(Q6_model)[4]

#Note, we don't need to consider effect of treatment or location because Berlin/Before are the referenc
boot::inv.logit(thresholds + gender_coef + genz_coef)
```

```
## Disagree|Neutral    Neutral|Agree
##       0.04099792       0.46313752
```

It is very unlikley for such a respondent to answer in the 'disagree' category; however, there is a much higher chance that they will answer neutral.

**

Step 3: Check model reliability

58

```
nominal_test(Q6_model)
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.


## Tests of nominal effects
##
## formula: response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location
##                 Df  logLik    AIC      LRT Pr(>Chi)
## <none>              -76.722 169.44
## Gender           1 -76.722 171.44 0.00007   0.9933
## Gen_Age_group    2 -76.260 172.52 0.92248   0.6305
## Location         1 -76.683 171.37 0.07769   0.7804
## Gender:Treatment
```

No evidence of violation for proportional odds assumption according to test with 'ordinal'.

Step 3b: Check that fitting model with alternative age group gives the same results.


```
##Create model
Q6_model_age <- clm(response_category ~ Gender + Gender:Treatment + Age_group + Location,
                    data = Q6_data, link = "logit")

summary(Q6_model_age)
```

```
## formula: response_category ~ Gender + Gender:Treatment + Age_group + Location
## data:    Q6_data
##
##  link  threshold nobs logLik AIC    niter max.grad cond.H
##  logit flexible  89   -74.57 165.15 6(0)  7.02e-13 8.7e+01
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## GenderMale                -0.8363     0.6490  -1.289   0.1975
## Age_group0-30              0.7133     0.7325   0.974   0.3302
## Age_group31-45             1.4681     0.7493   1.959   0.0501 .
## LocationKIEL               1.0922     0.8034   1.360   0.1740
## GenderFemale:TreatmentTRUE 0.2443    0.6213   0.393   0.6941
## GenderMale:TreatmentTRUE   0.3546    0.6592   0.538   0.5906
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Threshold coefficients:
##                 Estimate Std. Error z value
## Disagree|Neutral  -2.0949     0.8740  -2.397
## Neutral|Agree      1.0075     0.8322   1.211
```

Using different age category suggests stronger effects of Location and Age, but not significant.

Could we even analyse this as a continous variable with quadratic effect?

```
##Create model
Q6_model_agequad <- clm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q6_data, link = "logit")
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
summary(Q6_model_agequad)
```

```
## formula:
## response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location
## data:    Q6_data
##
##  link  threshold nobs logLik AIC    niter max.grad cond.H
##  logit flexible  89   -75.52 167.04 6(0)  2.89e-13 3.1e+02
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## GenderMale                -0.8027     0.6628  -1.211    0.226
## poly(Age, 2)1              2.9176     3.0753   0.949    0.343
## poly(Age, 2)2              3.2325     2.3131   1.397    0.162
## LocationKIEL              -0.3495     0.8689  -0.402    0.688
## GenderFemale:TreatmentTRUE 0.3049     0.6334   0.481    0.630
## GenderMale:TreatmentTRUE   0.8854     0.6411   1.381    0.167
##
## Threshold coefficients:
##                 Estimate Std. Error z value
## Disagree|Neutral  -2.9302     0.6736  -4.350
## Neutral|Agree      0.1063     0.5500   0.193
```

```
AIC(Q6_model, Q6_model_age, Q6_model_agequad)
```

```
##                  df      AIC
## Q6_model          8 169.4434
## Q6_model_age      8 165.1483
## Q6_model_agequad  8 167.0445
```

Again, using balanced age categories is most appropriate.

Step 4: Plot results

Plot raw data. Proportion of each group for each gender/treatment combo.

```
Q6_data %>%
  group_by(Gender, Treatment) %>%
  summarise(disagree = binom::binom.wilson(x = sum(response_category == "Disagree"), n = n()),
            neutral = binom::binom.wilson(x = sum(response_category == "Neutral"), n = n()),
            agree = binom::binom.wilson(x = sum(response_category == "Agree"), n = n())) %>%
  rowwise() %>%
  mutate(disagree_ymin = disagree$lower, disagree_ymax = disagree$upper, disagree_mean = disagree$mean,
         neutral_ymin = neutral$lower, neutral_ymax = neutral$upper, neutral_mean = neutral$mean,
         agree_ymin = agree$lower, agree_ymax = agree$upper, agree_mean = agree$mean) %>%
  ungroup() %>%
  tidyr::pivot_longer(cols = disagree_ymin:agree_mean) %>%
  tidyr::separate(col = name, into = c("category", "variable")) %>%
  tidyr::pivot_wider(names_from = variable, values_from = value) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = Treatment, ymin = ymin, ymax = ymax, group = category),
                    width = 0.25, position = position_dodge(width = 0.5)) +
      geom_point(aes(x = Treatment, y = mean, fill = category, group = category),
                 size = 3, position = position_dodge(width = 0.5), shape = 21) +
      scale_fill_manual(values = c("green", "red", "blue")) +
      facet_wrap(facets = ~Gender)}
```

`summarise()` has grouped output by 'Gender'. You can override using the `.groups` argument.

7. I would be willing to participate in collective actions to address microplastic pollution (e.g. demonstrations)

Step 1: Check whether data can be grouped for logistic regression.

```
ggplot() +
  geom_bar(data = clean_data, aes(x = Willingness_community_LK), colour = "black") +
  scale_x_continuous(name = "", limits = c(0.5, 7.5),
                     breaks = seq(1, 7, 1), labels = c("Strongly disagree", 2:6, "Strongly agree")) +
  scale_y_continuous(name = "Responses") +
  labs(title = "I would be willing to participate in collective actions to address microplastic pollutio
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))
```


Responses are more evenly spread than previous questions. Instead we will need to analyse using ordinal logistic regression (1-2; 3-5; 6-7).

Step 2: Fit preliminary model

We include an interaction between gender and treatment to see if effect of exhibition has different effects on men and women. To start with we use age groups defined by 'generations'. We will check to see if this differs if we use age groups that are more balanced. We include Location to make sure there is no difference between Kiel and Berlin.

```
Q7_data <- clean_data %>%
  dplyr::mutate(response_category = dplyr::case_when(Willingness_community_LK < 3 ~ "Disagree",
                                                     Willingness_community_LK >= 3 & Willingness_commun
                                                     Willingness_community_LK >= 6 ~ "Agree")) %>%
  #Make the new category ordered
  dplyr::mutate(response_category = ordered(response_category, levels = c("Disagree", "Neutral", "Agree
              across(.cols = c(Gen_Age_group, Location, Gender), .fns = as.factor))

##Create model
Q7_model <- clm(response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location,
              data = Q7_data, link = "logit")
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
summary(Q7_model)
```

```
## formula:
## response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location
## data:    Q7_data
##
##  link  threshold nobs logLik AIC     niter max.grad cond.H
##  logit flexible  89   -79.09 174.18 5(0)  4.92e-08 8.3e+01
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## GenderMale                -0.4873    0.6589  -0.739    0.460
## Gen_Age_groupGen_Z        -0.1523    0.7478  -0.204    0.839
## Gen_Age_groupMillenials    0.6168    0.6556   0.941    0.347
## LocationKIEL               1.3207    0.8167   1.617    0.106
## GenderFemale:TreatmentTRUE -0.9000   0.6432  -1.399    0.162
## GenderMale:TreatmentTRUE   -0.7283    0.6103  -1.193    0.233
##
## Threshold coefficients:
##                 Estimate Std. Error z value
## Disagree|Neutral -2.66502    0.84330  -3.160
## Neutral|Agree     0.09485    0.78037   0.122
```

*No significant effect of exhibition. If anything, suggests that exhibition reduced peoples willigness to partici-
pate.*

Step 3: Check model reliability

```
nominal_test(Q7_model)
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.


## Tests of nominal effects
##
## formula: response_category ~ Gender + Gender:Treatment + Gen_Age_group + Location
##                  Df  logLik   AIC     LRT Pr(>Chi)
## <none>              -79.092 174.18
## Gender           1 -78.625 175.25 0.9333   0.3340
## Gen_Age_group    2 -77.137 174.27 3.9111   0.1415
## Location         1 -78.821 175.64 0.5432   0.4611
## Gender:Treatment
```

No evidence of violation for proportional odds assumption according to test with 'ordinal'.

Step 3b: Check that fitting model with alternative age group gives the same results.

```
##Create model
Q7_model_age <- clm(response_category ~ Gender + Gender:Treatment + Age_group + Location,
                data = Q7_data, link = "logit")

summary(Q7_model_age)
```

```
## formula: response_category ~ Gender + Gender:Treatment + Age_group + Location
## data:    Q7_data
##
##  link  threshold nobs logLik AIC     niter max.grad cond.H
##  logit flexible  89   -73.55 163.10 6(0)  5.08e-13 9.0e+01
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## GenderMale                -0.1206     0.6608  -0.183 0.855141
## Age_group0-30              1.6100     0.7563   2.129 0.033276 *
## Age_group31-45             2.7583     0.8163   3.379 0.000728 ***
## LocationKIEL               2.5082     0.8825   2.842 0.004483 **
## GenderFemale:TreatmentTRUE -0.7484    0.6392  -1.171 0.241719
## GenderMale:TreatmentTRUE   -1.3850    0.6811  -2.033 0.042004 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##                 Estimate Std. Error z value
## Disagree|Neutral -1.3355     0.8445  -1.581
## Neutral|Agree     1.7595     0.8740   2.013
```

Using different age category changes our interpretation. We need to decide how to approach this!

Could we use quadratic response?

64

```
##Create model
Q7_model_agequad <- clm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q7_data, link = "logit")
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
summary(Q7_model_agequad)
```

```
## formula:
## response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location
## data:    Q7_data
##
##  link  threshold nobs logLik AIC    niter max.grad cond.H
##  logit flexible  89   -78.63 173.25 5(0)  8.86e-08 2.5e+02
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## GenderMale                  -0.5907     0.6643  -0.889   0.3739
## poly(Age, 2)1               -3.4167     2.7943  -1.223   0.2214
## poly(Age, 2)2               -3.2749     2.2549  -1.452   0.1464
## LocationKIEL                 1.9201     0.9117   2.106   0.0352 *
## GenderFemale:TreatmentTRUE  -0.8400     0.6447  -1.303   0.1926
## GenderMale:TreatmentTRUE    -0.7384     0.6224  -1.186   0.2355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##                 Estimate Std. Error z value
## Disagree|Neutral  -2.9242     0.6599  -4.431
## Neutral|Agree     -0.1298     0.5508  -0.236
```

```
AIC(Q7_model, Q7_model_age, Q7_model_agequad)
```

```
##                  df      AIC
## Q7_model          8 174.1842
## Q7_model_age      8 163.1014
## Q7_model_agequad  8 173.2530
```

Again, balanced age categories are more appropriate.

Step 4: Plot results

Plot raw data. Proportion of each group for each gender/treatment combo.

```
Q7_data %>%
  group_by(Gender, Treatment) %>%
  summarise(disagree = binom::binom.wilson(x = sum(response_category == "Disagree"), n = n()),
            neutral = binom::binom.wilson(x = sum(response_category == "Neutral"), n = n()),
            agree = binom::binom.wilson(x = sum(response_category == "Agree"), n = n())) %>%
```

65

```
##Create model
Q7_model_agequad <- clm(response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                        data = Q7_data, link = "logit")
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
summary(Q7_model_agequad)
```

```
## formula:
## response_category ~ Gender + Gender:Treatment + poly(Age, 2) + Location
## data:    Q7_data
##
##  link  threshold nobs logLik AIC    niter max.grad cond.H
##  logit flexible  89   -78.63 173.25 5(0)  8.86e-08 2.5e+02
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## GenderMale                  -0.5907     0.6643  -0.889   0.3739
## poly(Age, 2)1               -3.4167     2.7943  -1.223   0.2214
## poly(Age, 2)2               -3.2749     2.2549  -1.452   0.1464
## LocationKIEL                 1.9201     0.9117   2.106   0.0352 *
## GenderFemale:TreatmentTRUE  -0.8400     0.6447  -1.303   0.1926
## GenderMale:TreatmentTRUE    -0.7384     0.6224  -1.186   0.2355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##                 Estimate Std. Error z value
## Disagree|Neutral  -2.9242     0.6599  -4.431
## Neutral|Agree     -0.1298     0.5508  -0.236
```

```
AIC(Q7_model, Q7_model_age, Q7_model_agequad)
```

```
##                  df      AIC
## Q7_model          8 174.1842
## Q7_model_age      8 163.1014
## Q7_model_agequad  8 173.2530
```

Again, balanced age categories are more appropriate.

Step 4: Plot results

Plot raw data. Proportion of each group for each gender/treatment combo.
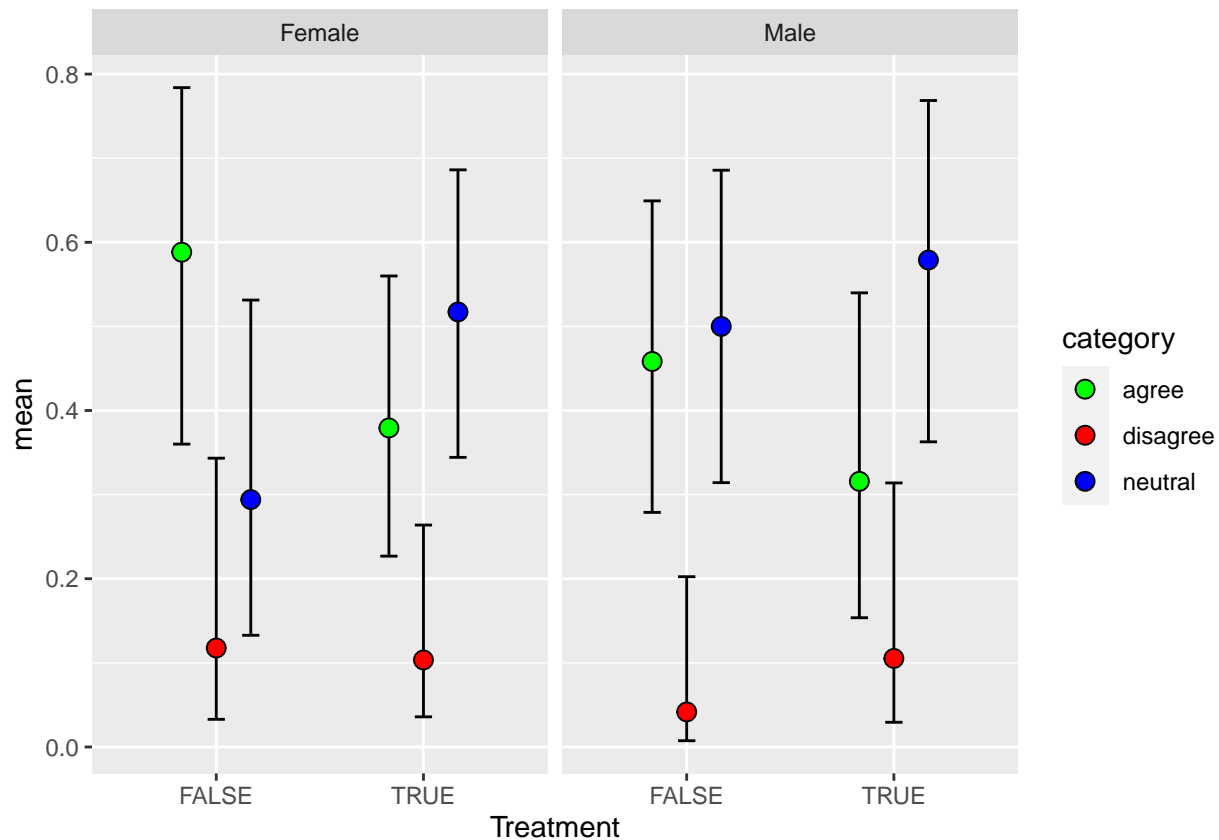
```
Q7_data %>%
  group_by(Gender, Treatment) %>%
  summarise(disagree = binom::binom.wilson(x = sum(response_category == "Disagree"), n = n()),
            neutral = binom::binom.wilson(x = sum(response_category == "Neutral"), n = n()),
            agree = binom::binom.wilson(x = sum(response_category == "Agree"), n = n())) %>%
```

```
rowwise() %>%
mutate(disagree_ymin = disagree$lower, disagree_ymax = disagree$upper, disagree_mean = disagree$mean,
       neutral_ymin = neutral$lower, neutral_ymax = neutral$upper, neutral_mean = neutral$mean,
       agree_ymin = agree$lower, agree_ymax = agree$upper, agree_mean = agree$mean) %>%
ungroup() %>%
tidyr::pivot_longer(cols = disagree_ymin:agree_mean) %>%
tidyr::separate(col = name, into = c("category", "variable")) %>%
tidyr::pivot_wider(names_from = variable, values_from = value) %>%
{ggplot(.) +
    geom_errorbar(aes(x = Treatment, ymin = ymin, ymax = ymax, group = category),
                  width = 0.25, position = position_dodge(width = 0.5)) +
    geom_point(aes(x = Treatment, y = mean, fill = category, group = category),
               size = 3, position = position_dodge(width = 0.5), shape = 21) +
    scale_fill_manual(values = c("green", "red", "blue")) +
    facet_wrap(facets = ~Gender)}
```

`summarise()` has grouped output by 'Gender'. You can override using the `.groups` argument.

8. Are people willing to sign petition or sign up for newsletter?

Step 1: How should we organise the data?

People can choose both of sign petition and sign-up for newsletter. We could analyse any activity v. no activity or analyse each type of activity separately using an 'activity type' variable. We will try both.

Step 1a: Any activity v. no activity

```
Q8a_data <- clean_data %>%
  dplyr::mutate(across(.cols = c(Newsletter, Petition), .fns = ~tidyr::replace_na(., 0L))) %>%
  rowwise() %>%
  dplyr::mutate(any_action = as.numeric(sum(Newsletter, Petition) > 0))
```

```
Q8a_model <- glm(any_action ~ Gender + Gender:Treatment + Gen_Age_group + Location,
                 family = binomial(link = "logit"), data = Q8a_data)
```

```
summary(Q8a_model)
```

```
##
## Call:
## glm(formula = any_action ~ Gender + Gender:Treatment + Gen_Age_group +
##     Location, family = binomial(link = "logit"), data = Q8a_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0715   0.4227   0.6732   0.7847   1.2196
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 0.01116    0.79742   0.014   0.9888
## GenderMale                  0.34922    0.79516   0.439   0.6605
## Gen_Age_groupGen_Z          2.01001    0.95712   2.100   0.0357 *
## Gen_Age_groupMillenials     1.00890    0.66752   1.511   0.1307
## LocationKIEL                1.15347    0.86254   1.337   0.1811
## GenderFemale:TreatmentTRUE -0.10982    0.74248  -0.148   0.8824
## GenderMale:TreatmentTRUE   -0.14774    0.74998  -0.197   0.8438
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 99.543  on 88  degrees of freedom
## Residual deviance: 94.013  on 82  degrees of freedom
## AIC: 108.01
##
## Number of Fisher Scoring iterations: 4
```

If we look at chance to take any action GenZ are most likely to take action (significantly more likely than GenX+). Millenials are also more likely, but not significant. Treatment has no effect.

Step 1b: Separate activities

```
Q8b_data <- clean_data %>%
  mutate(across(.cols = c(Newsletter, Petition), .fns = ~tidyr::replace_na(., 0L))) %>%
  tidyr::pivot_longer(cols = c(Newsletter, Petition), names_to = "Action")
```

```
Q8b_model <- glmer(value ~ Gender + Gender:Treatment + Gen_Age_group +
                    Location + Action + Action:Gen_Age_group + (1|X_id), family = binomial(link = "log:
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(Q8b_model)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: value ~ Gender + Gender:Treatment + Gen_Age_group + Location +
##     Action + Action:Gen_Age_group + (1 | X_id)
##    Data: Q8b_data
##
##      AIC      BIC   logLik deviance df.resid
##    199.1    234.1    -88.6    177.1      167
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7414 -0.5188 -0.2713  0.6627  4.2968
##
## Random effects:
##  Groups Name        Variance  Std.Dev.
##  X_id   (Intercept) 1.971e-14 1.404e-07
## Number of obs: 178, groups:  X_id, 89
##
## Fixed effects:
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -2.16572    0.78533  -2.758  0.00582 **
## GenderMale                           0.37426    0.57809   0.647  0.51737
## Gen_Age_groupGen_Z                  -0.81727    1.22413  -0.668  0.50437
## Gen_Age_groupMillenials              0.48862    0.76087   0.642  0.52076
## LocationKIEL                         0.82362    0.63919   1.289  0.19756
## ActionPetition                       1.93276    0.65129   2.968  0.00300 **
## GenderFemale:TreatmentTRUE           0.06725    0.56393   0.119  0.90508
## GenderMale:TreatmentTRUE            -0.14959    0.52755  -0.284  0.77675
## Gen_Age_groupGen_Z:ActionPetition    2.99993    1.43126   2.096  0.03608 *
## Gen_Age_groupMillenials:ActionPetition 0.34264  0.82584   0.415  0.67821
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) GndrMl Gn_A_G_Z Gn_A_M LcKIEL ActnPt GF:TTR GM:TTR G_A_G_Z:
## GenderMale  -0.499
## Gn_Ag_grG_Z -0.481  0.053
## Gn_Ag_grpMl -0.710  0.028  0.457
## LocatinKIEL -0.555  0.030  0.343    0.539
## ActionPettn -0.592  0.034  0.372    0.598  0.161
## GndrF:TTRUE -0.462  0.629  0.035   -0.041  0.044  0.010
## GndrM:TTRUE -0.012 -0.409  0.056   -0.012  0.055 -0.017  0.019
## Gn_A_G_Z:AP  0.253  0.021 -0.772   -0.270 -0.072 -0.454  0.010 -0.009
## Gn_Ag_gM:AP  0.457 -0.004 -0.292   -0.713 -0.126 -0.788 -0.007  0.004  0.358
```

68

```
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Significantly more likely to sign a petition than sign up for a newsletter! Effect of age seems to be because GenZ are much more likely to sign petitions!

Step 2: Check model reliability

Use DHARMa to check for issues of heteroskedasticity or outliers.

```
##Check model using DHARMa
simulationOutput <- simulateResiduals(Q8a_model, seed = 123, n = 5000)
plot(simulationOutput)
```
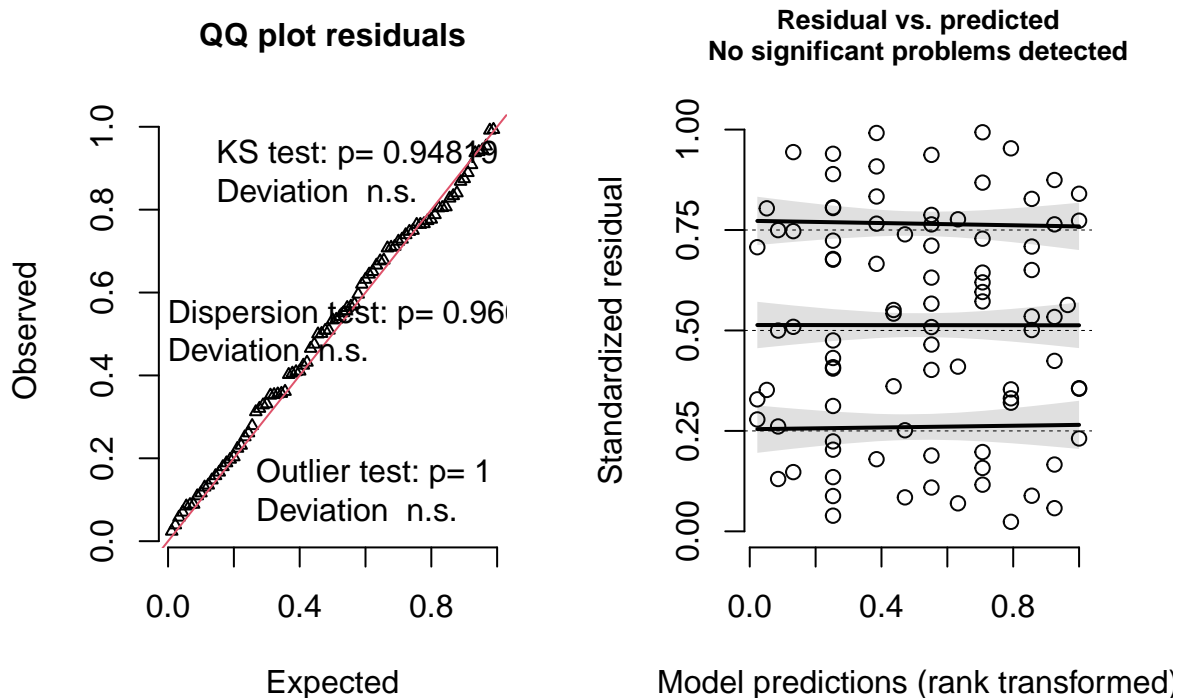
DHARMa residual diagnostics


```
#Check there are no unusual residual patterns for each predictor
plotResiduals(simulationOutput, form = as.factor(Q8a_data$Gender))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
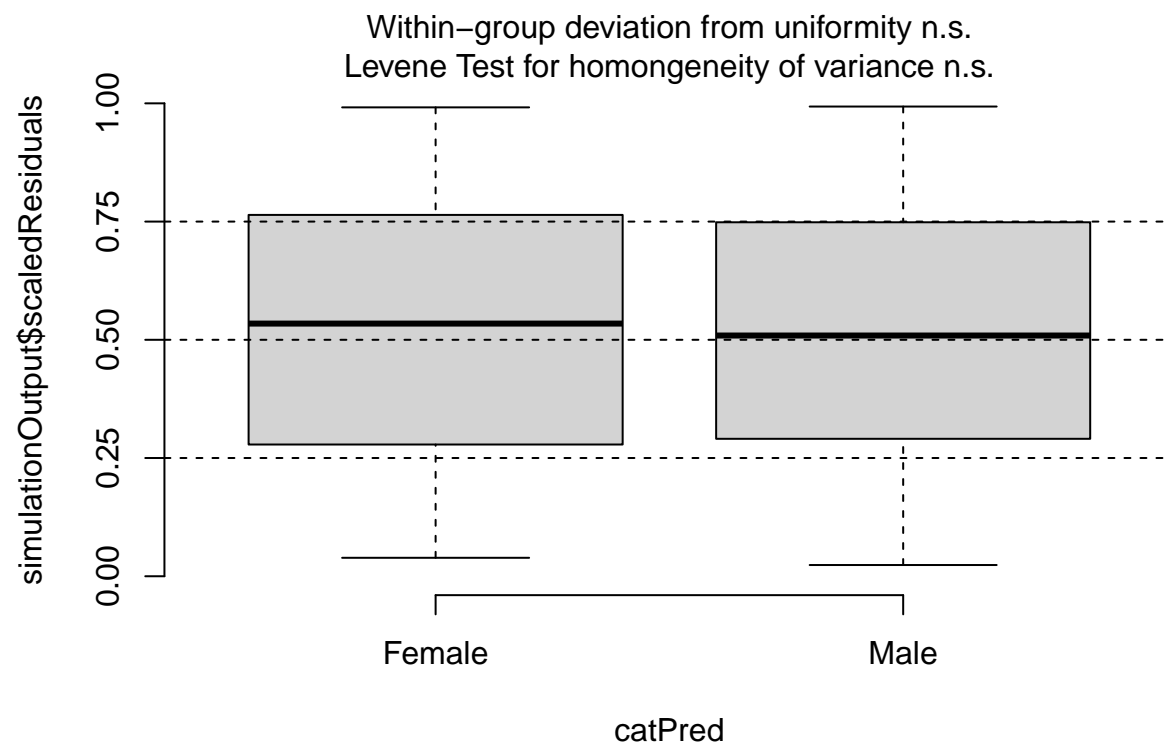
```
plotResiduals(simulationOutput, form = as.factor(Q8a_data$Age_group))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.

```
plotResiduals(simulationOutput, form = as.factor(Q8a_data$Treatment))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.

```
plotResiduals(simulationOutput, form = as.factor(Q8a_data$Location))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.

```
##Check model using DHARMa
simulationOutput <- simulateResiduals(Q8b_model, seed = 123, n = 5000)
plot(simulationOutput)
```
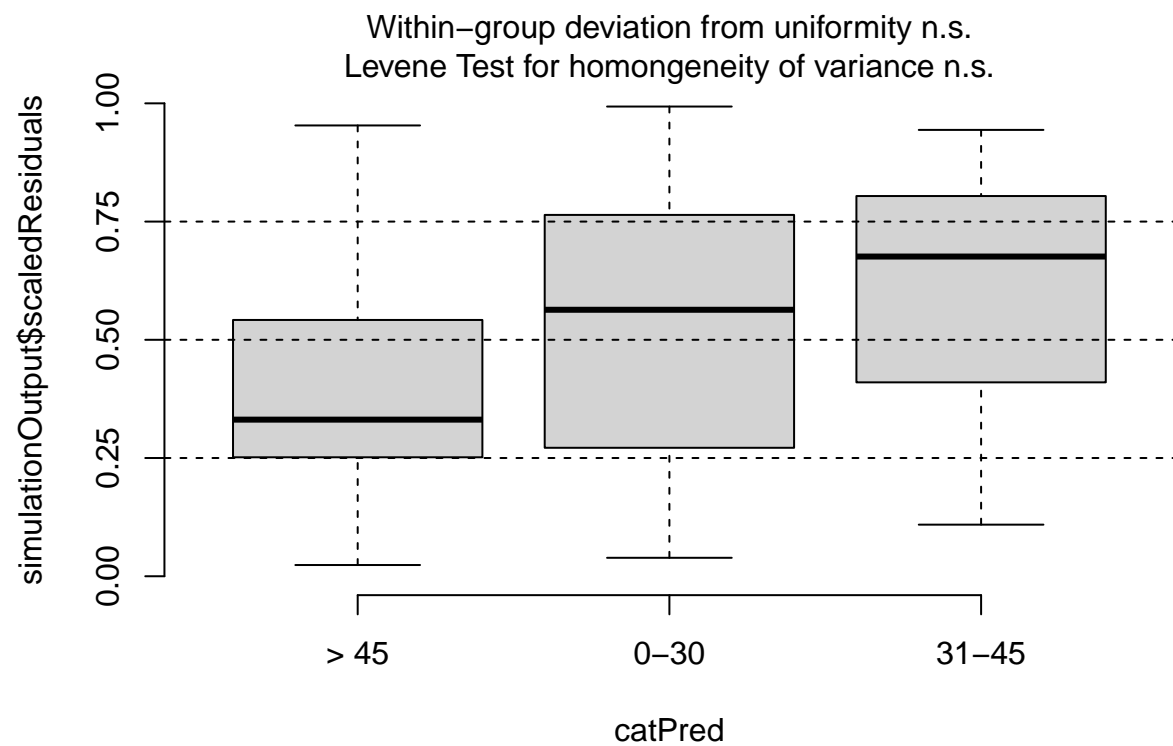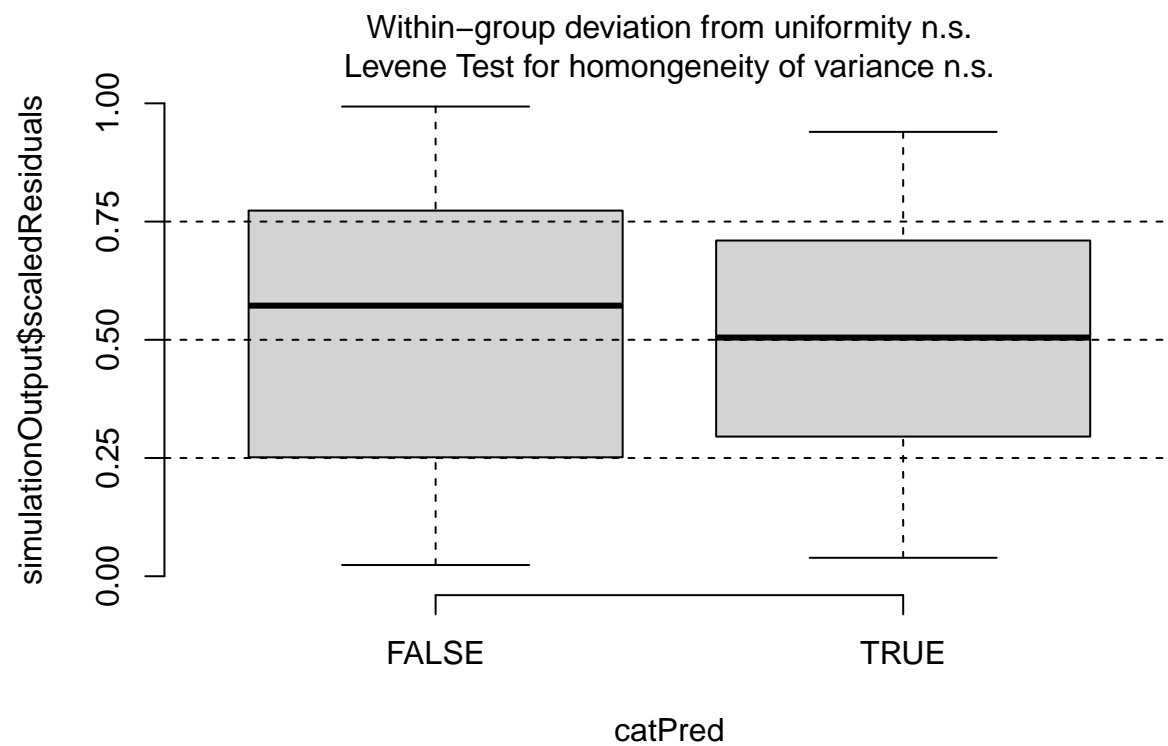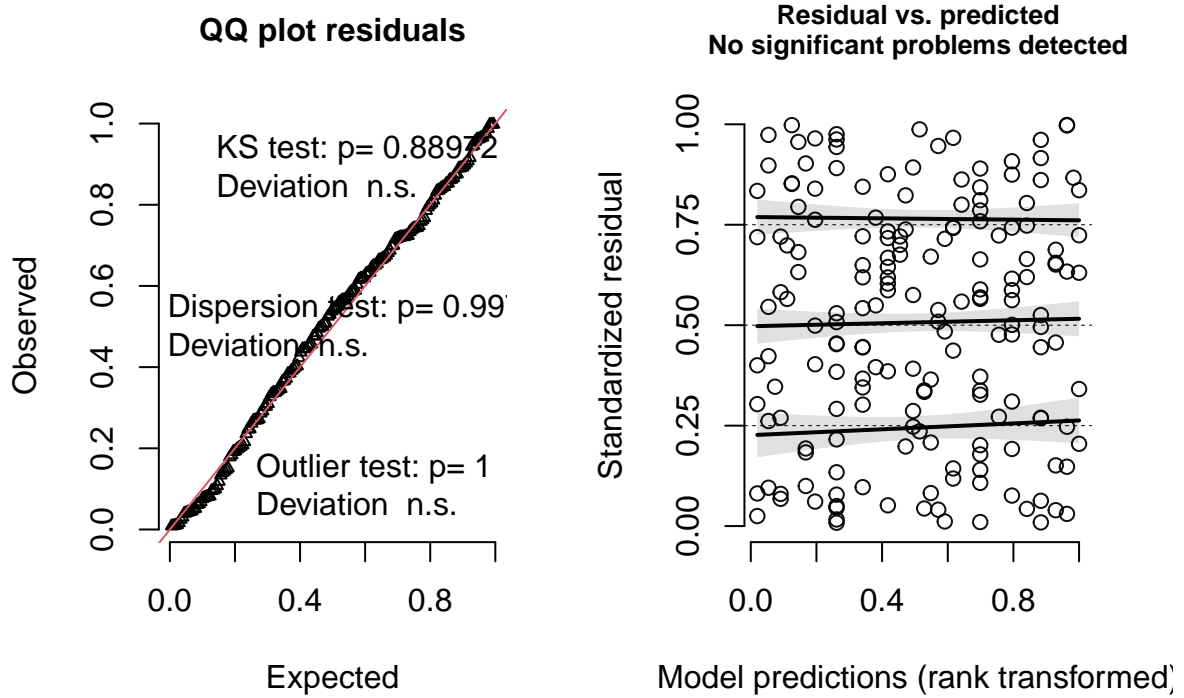
DHARMa residual diagnostics

QQ plot residuals

KS test: p= 0.88972
Deviation n.s.

Dispersion test: p= 0.99
Deviation n.s.

Outlier test: p= 1
Deviation n.s.

Residual vs. predicted
No significant problems detected


```
#Check there are no unusual residual patterns for each predictor
plotResiduals(simulationOutput, form = as.factor(Q8b_data$Gender))
```

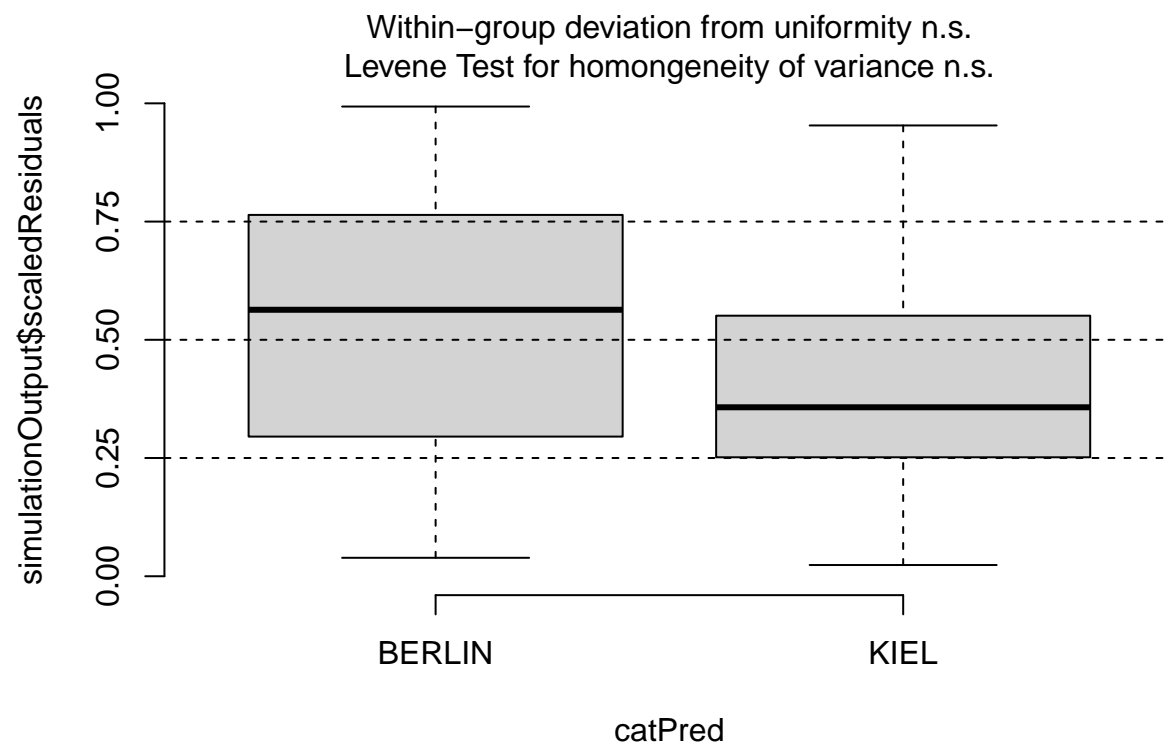Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
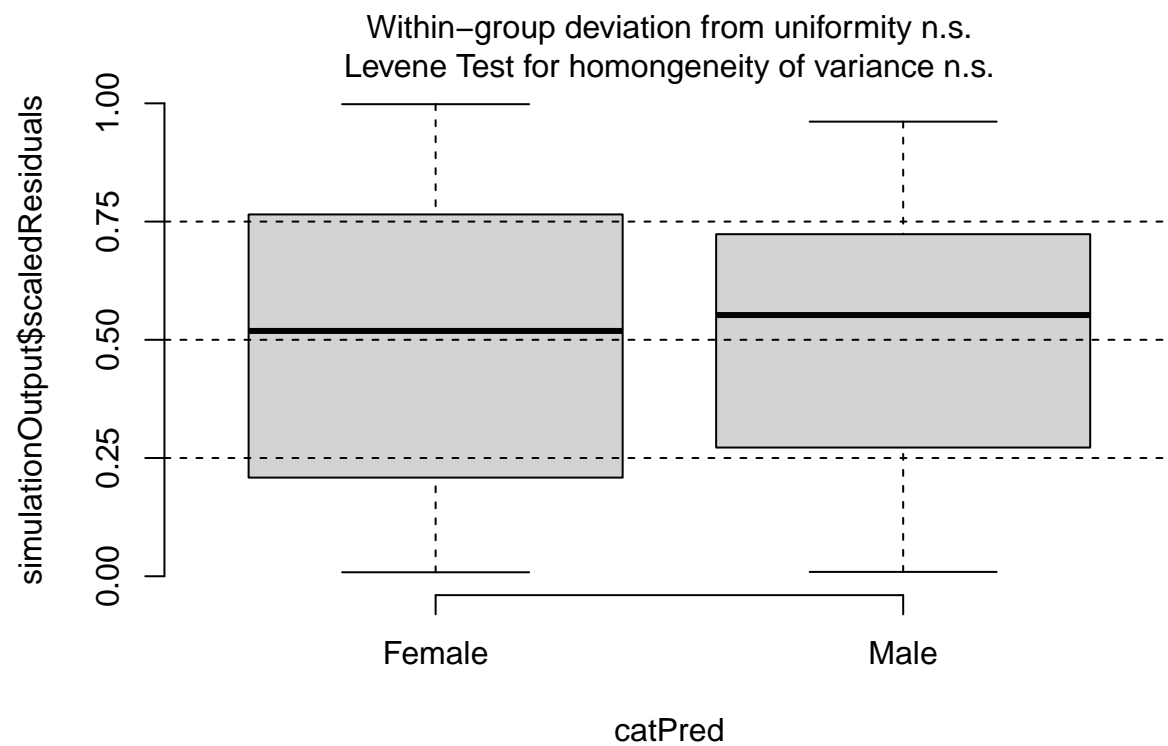
```
plotResiduals(simulationOutput, form = as.factor(Q8b_data$Age_group))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
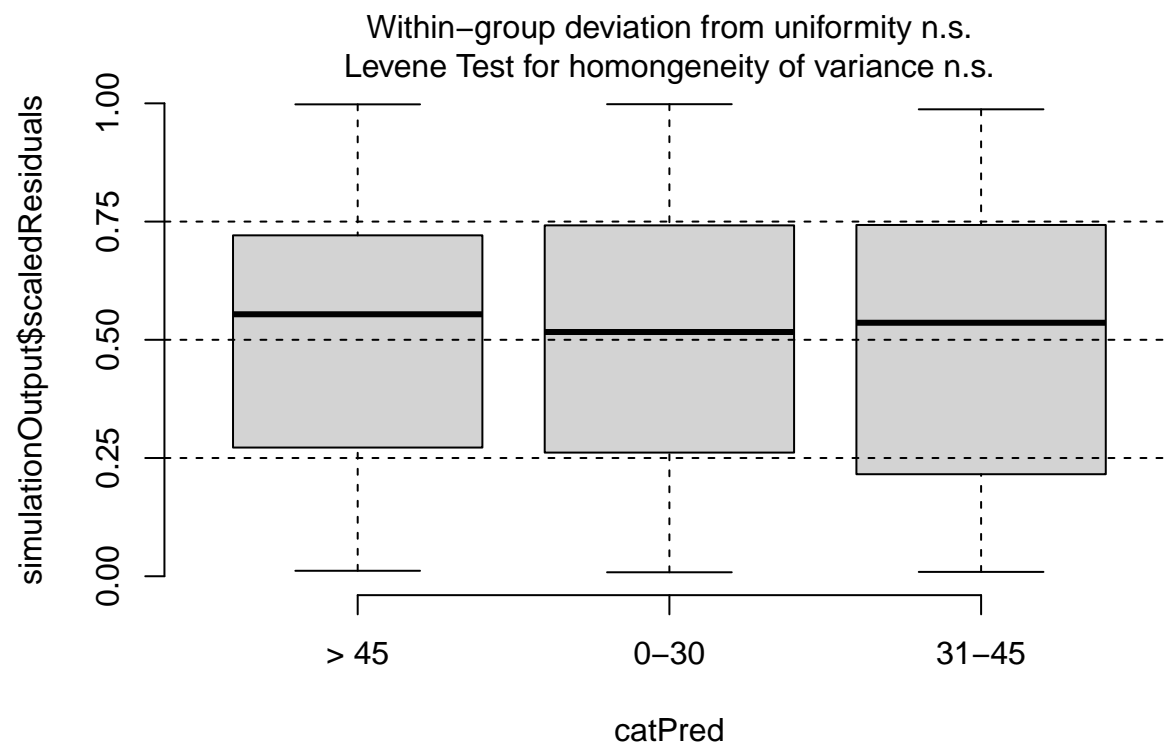
```
plotResiduals(simulationOutput, form = as.factor(Q8b_data$Treatment))
```

Within−group deviation from uniformity n.s.
Levene Test for homongeneity of variance n.s.
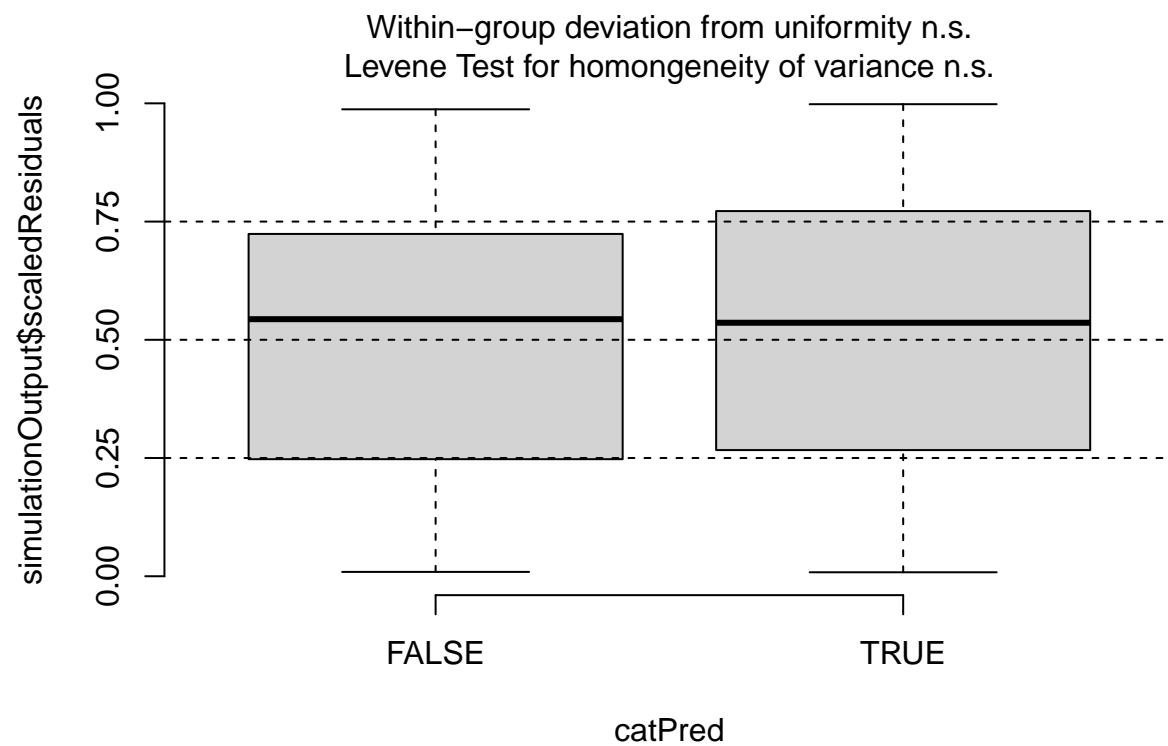
```
plotResiduals(simulationOutput, form = as.factor(Q8b_data$Location))
```

Within−group deviation from uniformity n.s.
Levene Test for homogeneity of variance significant
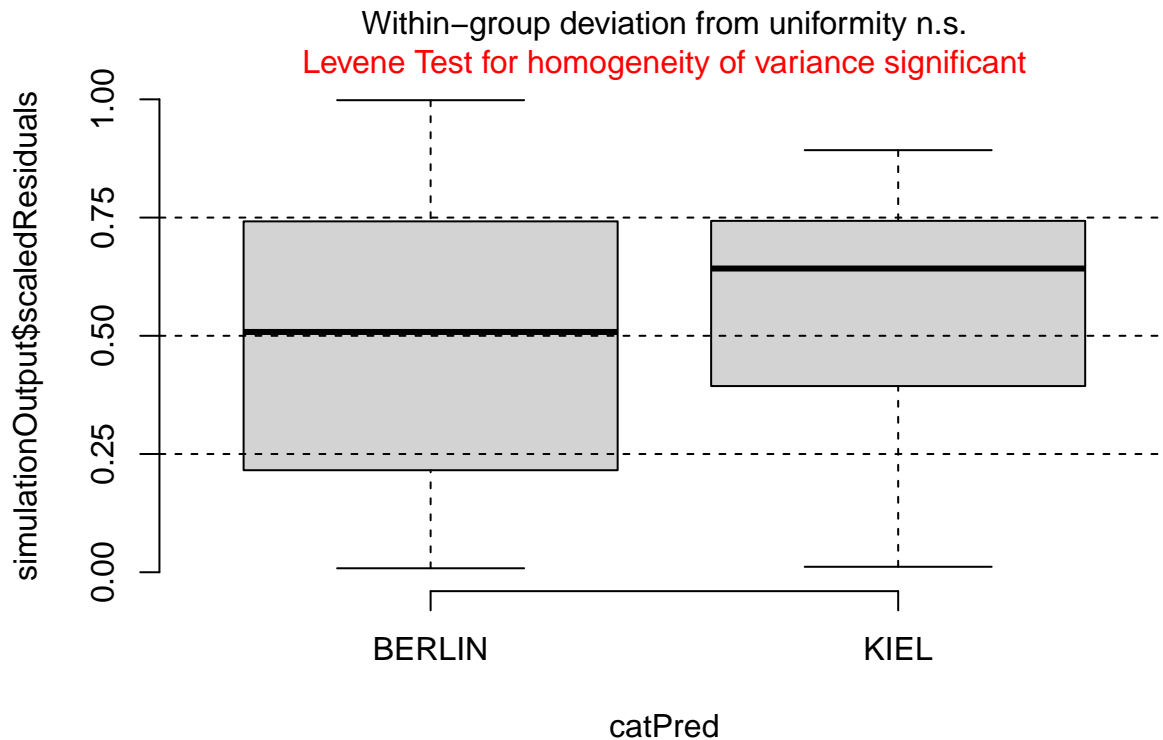
No apparent issues with model.

Step 2b: Check that fitting model with alternative age group gives the same results.

```
##Create model
Q8a_model_age <- Q8a_model <- glm(any_action ~ Gender + Gender:Treatment + Age_group + Location,
                family = binomial(link = "logit"), data = Q8a_data)

summary(Q8a_model_age)
```

```
##
## Call:
## glm(formula = any_action ~ Gender + Gender:Treatment + Age_group +
##     Location, family = binomial(link = "logit"), data = Q8a_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0677   0.2898   0.6462   0.6841   1.5848
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -0.7665     0.9199  -0.833  0.40467
## GenderMale               0.3979     0.8096   0.492  0.62305
## Age_group0-30            2.1015     0.8171   2.572  0.01011 *
```

```
## Age_group31-45              2.3809      0.8785   2.710  0.00673 **
## LocationKIEL                1.8156      0.9562   1.899  0.05759 .
## GenderFemale:TreatmentTRUE  -0.1541     0.7445  -0.207  0.83608
## GenderMale:TreatmentTRUE    -0.6790     0.8304  -0.818  0.41354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 99.543  on 88  degrees of freedom
## Residual deviance: 89.688  on 82  degrees of freedom
## AIC: 103.69
##
## Number of Fisher Scoring iterations: 4
```

Stronger effect of age (~Millenials are also significant)

```
Q8a_data %>%
  mutate(age_cutoff = cut(Age, breaks = c(seq(15, 45, 5), 85)),
         age_cutoff_numeric = as.numeric(age_cutoff)) %>%
  group_by(age_cutoff_numeric) %>%
  summarise(binom::binom.wilson(x = sum(any_action), n = n()),
            age_cutoff = first(age_cutoff),
            n = n()) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = age_cutoff_numeric, ymin = lower, ymax = upper), width = 0.25) +
      geom_point(aes(x = age_cutoff_numeric, y = mean), size = 3, shape = 21, fill = "grey45") +
      geom_text(aes(x = age_cutoff_numeric + 0.15, y = mean, label = n), size = 4) +
      scale_x_continuous(breaks = seq(1:max(.$age_cutoff_numeric)),
                         labels = .$age_cutoff) +
      scale_y_continuous(limits = c(0, 1)) +
      theme_classic()}
```

Can we fit this as quadratic?

```
##Create model
Q8a_model_agequad <- Q8a_model <- glm(any_action ~ Gender + Gender:Treatment + poly(Age, 2) + Location,
                family = binomial(link = "logit"), data = Q8a_data)

summary(Q8a_model_agequad)
```
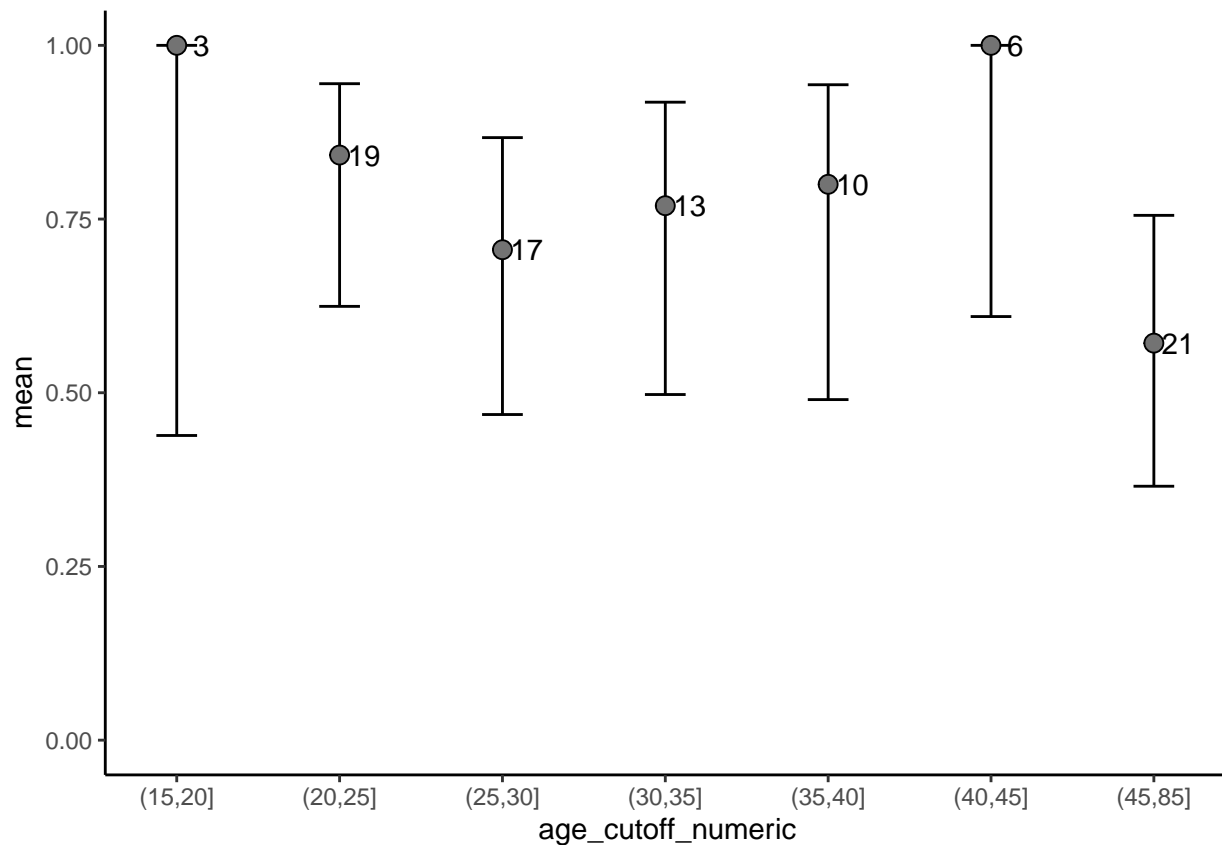
```
##
## Call:
## glm(formula = any_action ~ Gender + Gender:Treatment + poly(Age,
##     2) + Location, family = binomial(link = "logit"), data = Q8a_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9938   0.4036   0.5887   0.8148   1.2186
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              0.773485   0.622472   1.243   0.2140
## GenderMale               0.459139   0.813147   0.565   0.5723
## poly(Age, 2)1           -6.624517   3.268762  -2.027   0.0427 *
## poly(Age, 2)2            3.681843   2.712832   1.357   0.1747
## LocationKIEL             1.257716   0.951607   1.322   0.1863
## GenderFemale:TreatmentTRUE  0.087408   0.750924   0.116   0.9073
## GenderMale:TreatmentTRUE   -0.007741   0.758240  -0.010   0.9919
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 99.543  on 88  degrees of freedom
## Residual deviance: 92.546  on 82  degrees of freedom
## AIC: 106.55
##
## Number of Fisher Scoring iterations: 4
```

```
AIC(Q8a_model, Q8a_model_age, Q8a_model_agequad)
```

```
##                   df      AIC
## Q8a_model          7 106.5460
## Q8a_model_age      7 103.6877
## Q8a_model_agequad  7 106.5460
```

Create model
```
Q8b_model_age <- glmer(value ~ Gender + Gender:Treatment + Age_group +
                    Location + Action + Action:Gen_Age_group + (1|X_id), family = binomial(link = "log
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(Q8b_model_age)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: value ~ Gender + Gender:Treatment + Age_group + Location + Action +
##     Action:Gen_Age_group + (1 | X_id)
##    Data: Q8b_data
##
##      AIC      BIC   logLik deviance df.resid
##    197.8    239.2    -85.9    171.8      165
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8024 -0.5001 -0.2360  0.6305  4.2369
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  X_id   (Intercept) 0        0
## Number of obs: 178, groups:  X_id, 89
##
## Fixed effects:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -2.9019     0.9211  -3.150  0.00163
## GenderMale                        0.4967     0.5910   0.840  0.40068
## Age_group0-30                     1.4605     0.9923   1.472  0.14107
## Age_group31-45                    1.8643     0.8810   2.116  0.03433
## LocationKIEL                      1.1940     0.7036   1.697  0.08970
```

```
## ActionPetition                                    2.1336     0.7047    3.028   0.00247
## GenderFemale:TreatmentTRUE                         0.1474     0.5736    0.257   0.79720
## GenderMale:TreatmentTRUE                          -0.5850     0.5796   -1.009   0.31283
## ActionNewsletter:Gen_Age_groupGen_Z              -1.5937     1.4031   -1.136   0.25605
## ActionPetition:Gen_Age_groupGen_Z                 1.2214     1.2276    0.995   0.31976
## ActionNewsletter:Gen_Age_groupMillenials         -0.4410     0.9213   -0.479   0.63216
## ActionPetition:Gen_Age_groupMillenials           -0.2777     0.9172   -0.303   0.76204
##
## (Intercept)                                 **
## GenderMale
## Age_group0-30
## Age_group31-45                              *
## LocationKIEL                                .
## ActionPetition                              **
## GenderFemale:TreatmentTRUE
## GenderMale:TreatmentTRUE
## ActionNewsletter:Gen_Age_groupGen_Z
## ActionPetition:Gen_Age_groupGen_Z
## ActionNewsletter:Gen_Age_groupMillenials
## ActionPetition:Gen_Age_groupMillenials
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) GndrMl A_0-30 A_31-4 LcKIEL ActnPt GF:TTR GM:TTR AN:G_A_G
## GenderMale  -0.460
## Age_grp0-30 -0.415  0.024
## Ag_grp31-45 -0.479  0.082  0.854
## LocatinKIEL -0.610  0.043  0.273  0.310
## ActionPettn -0.623  0.048  0.232  0.267  0.203
## GndrF:TTRUE -0.417  0.625  0.070  0.072  0.050  0.015
## GndrM:TTRUE  0.100 -0.421 -0.127 -0.282 -0.024 -0.046  0.000
## AcN:G_A_G_Z -0.232  0.042 -0.446 -0.308  0.194  0.236 -0.012  0.078
## AcP:G_A_G_Z  0.068  0.077 -0.640 -0.498  0.107 -0.301  0.005  0.079  0.393
## ActnN:G_A_M -0.316 -0.001 -0.520 -0.464  0.292  0.358 -0.092  0.102  0.586
## ActnP:G_A_M  0.152 -0.010 -0.716 -0.662  0.138 -0.406 -0.096  0.099  0.418
##            AP:G_A_G AN:G_A_M
## GenderMale
## Age_grp0-30
## Ag_grp31-45
## LocatinKIEL
## ActionPettn
## GndrF:TTRUE
## GndrM:TTRUE
## AcN:G_A_G_Z
## AcP:G_A_G_Z
## ActnN:G_A_M  0.463
## ActnP:G_A_M  0.711     0.552
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

AGAIN! We get different interpretations. . . not sure what the best way to deal with these

```
Q8b_data %>%
  mutate(age_cutoff = cut(Age, breaks = c(seq(15, 45, 5), 85)),
         age_cutoff_numeric = as.numeric(age_cutoff)) %>%
  group_by(age_cutoff_numeric) %>%
  summarise(binom::binom.wilson(x = sum(value), n = n()),
            age_cutoff = first(age_cutoff),
            n = n()) %>%
  {ggplot(.) +
      geom_errorbar(aes(x = age_cutoff_numeric, ymin = lower, ymax = upper), width = 0.25) +
      geom_point(aes(x = age_cutoff_numeric, y = mean), size = 3, shape = 21, fill = "grey45") +
      geom_text(aes(x = age_cutoff_numeric + 0.15, y = mean, label = n), size = 4) +
      scale_x_continuous(breaks = seq(1:max(.$age_cutoff_numeric)),
                         labels = .$age_cutoff) +
      scale_y_continuous(limits = c(0, 1)) +
      theme_classic()}
```
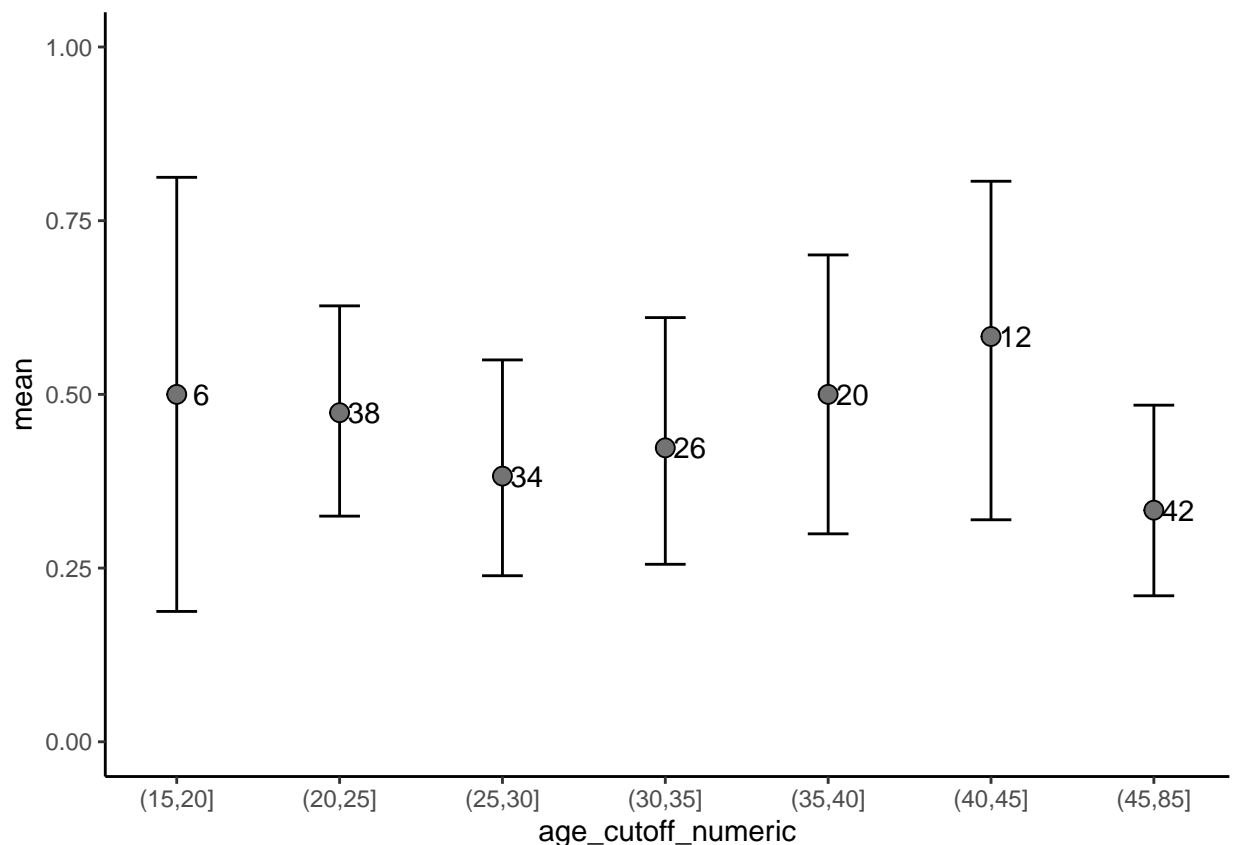

Can we fit this as quadratic?

```
##Create model
Q8b_model_agequad <- glmer(value ~ Gender + Gender:Treatment + poly(Age, 2) +
                    Location + Action + Action:Gen_Age_group + (1|X_id), family = binomial(link = "log
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(Q8b_model_agequad)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: binomial  ( logit )
## Formula: value ~ Gender + Gender:Treatment + poly(Age, 2) + Location +
##     Action + Action:Gen_Age_group + (1 | X_id)
##    Data: Q8b_data
##
##      AIC      BIC   logLik deviance df.resid
##    201.8    243.1    -87.9    175.8      165
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7415 -0.5080 -0.2583  0.6503  4.0811
##
## Random effects:
##  Groups Name        Variance  Std.Dev.
##  X_id   (Intercept) 3.397e-18 1.843e-09
## Number of obs: 178, groups:  X_id, 89
##
## Fixed effects:
##                                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)                            -1.66079    0.92370  -1.798   0.0722
## GenderMale                              0.35905    0.58714   0.612   0.5409
## poly(Age, 2)1                          -7.20477    6.45913  -1.115   0.2647
## poly(Age, 2)2                           0.99834    3.62057   0.276   0.7827
## LocationKIEL                            1.11577    0.71885   1.552   0.1206
## ActionPetition                          1.96938    0.66121   2.978   0.0029
## GenderFemale:TreatmentTRUE              0.15361    0.58025   0.265   0.7912
## GenderMale:TreatmentTRUE               -0.10370    0.53919  -0.192   0.8475
## ActionNewsletter:Gen_Age_groupGen_Z    -1.97252    1.73329  -1.138   0.2551
## ActionPetition:Gen_Age_groupGen_Z       0.98764    1.53314   0.644   0.5194
## ActionNewsletter:Gen_Age_groupMillenials -0.24154  1.03791  -0.233   0.8160
## ActionPetition:Gen_Age_groupMillenials  0.08033    0.91954   0.087   0.9304
##
## (Intercept)                            .
## GenderMale
## poly(Age, 2)1
## poly(Age, 2)2
## LocationKIEL
## ActionPetition                         **
## GenderFemale:TreatmentTRUE
## GenderMale:TreatmentTRUE
## ActionNewsletter:Gen_Age_groupGen_Z
## ActionPetition:Gen_Age_groupGen_Z
## ActionNewsletter:Gen_Age_groupMillenials
## ActionPetition:Gen_Age_groupMillenials
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) GndrMl p(A,2)1 p(A,2)2 LcKIEL ActnPt GF:TTR GM:TTR AN:G_A_G
```

```
## GenderMale  -0.385
## poly(Ag,2)1 -0.498 -0.032
## poly(Ag,2)2  0.361  0.156 -0.504
## LocatinKIEL -0.331 -0.024 -0.277  -0.158
## ActionPettn -0.492  0.027 -0.062  -0.047   0.194
## GndrF:TTRUE -0.277  0.631 -0.172   0.201   0.030  0.013
## GndrM:TTRUE  0.056 -0.368 -0.116   0.185   0.033 -0.001  0.057
## AcN:G_A_G_Z -0.656 -0.021  0.659  -0.549   0.123  0.247 -0.129 -0.071
## AcP:G_A_G_Z -0.538 -0.005  0.766  -0.589   0.055 -0.151 -0.128 -0.089  0.620
## ActnN:G_A_M -0.800 -0.020  0.658  -0.455   0.226  0.412 -0.165 -0.091  0.712
## ActnP:G_A_M -0.530 -0.019  0.746  -0.437   0.119 -0.253 -0.186 -0.112  0.593
##             AP:G_A_G AN:G_A_M
## GenderMale
## poly(Ag,2)1
## poly(Ag,2)2
## LocatinKIEL
## ActionPettn
## GndrF:TTRUE
## GndrM:TTRUE
## AcN:G_A_G_Z
## AcP:G_A_G_Z
## ActnN:G_A_M  0.623
## ActnP:G_A_M  0.774     0.642
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
AIC(Q8b_model, Q8b_model_age, Q8b_model_agequad)
```

```
##                 df      AIC
## Q8b_model       11 199.1407
## Q8b_model_age   13 197.8221
## Q8b_model_agequad 13 201.7615
```

Again, age categories are more appropriate.

Step 3: Plot results

```
## Plot results (gender and treatment)
plot_data <- Q8a_data %>%
  dplyr::group_by(Gender, Treatment) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(any_action), n = n()), .groups = "drop")

gender <- ggplot() +
  geom_errorbar(data = plot_data, aes(x = Treatment, ymin = lower, ymax = upper, group = Gender),
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data, aes(x = Treatment, y = mean, fill = Gender),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "I would be willing to sign a petition/sign up for newsletter") +
```

```r
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))

## Plot results (gender and treatment)
plot_data_age <- Q8a_data %>%
  dplyr::group_by(Gen_Age_group, Treatment) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(any_action), n = n()), .groups = "drop")

age <- ggplot() +
  geom_errorbar(data = plot_data_age, aes(x = Treatment, ymin = lower, ymax = upper, group = Gen_Age_gro
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data_age, aes(x = Treatment, y = mean, fill = Gen_Age_group),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "I would be willing to sign a petition/sign up for newsletter") +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))

library(patchwork)
gender + age
```
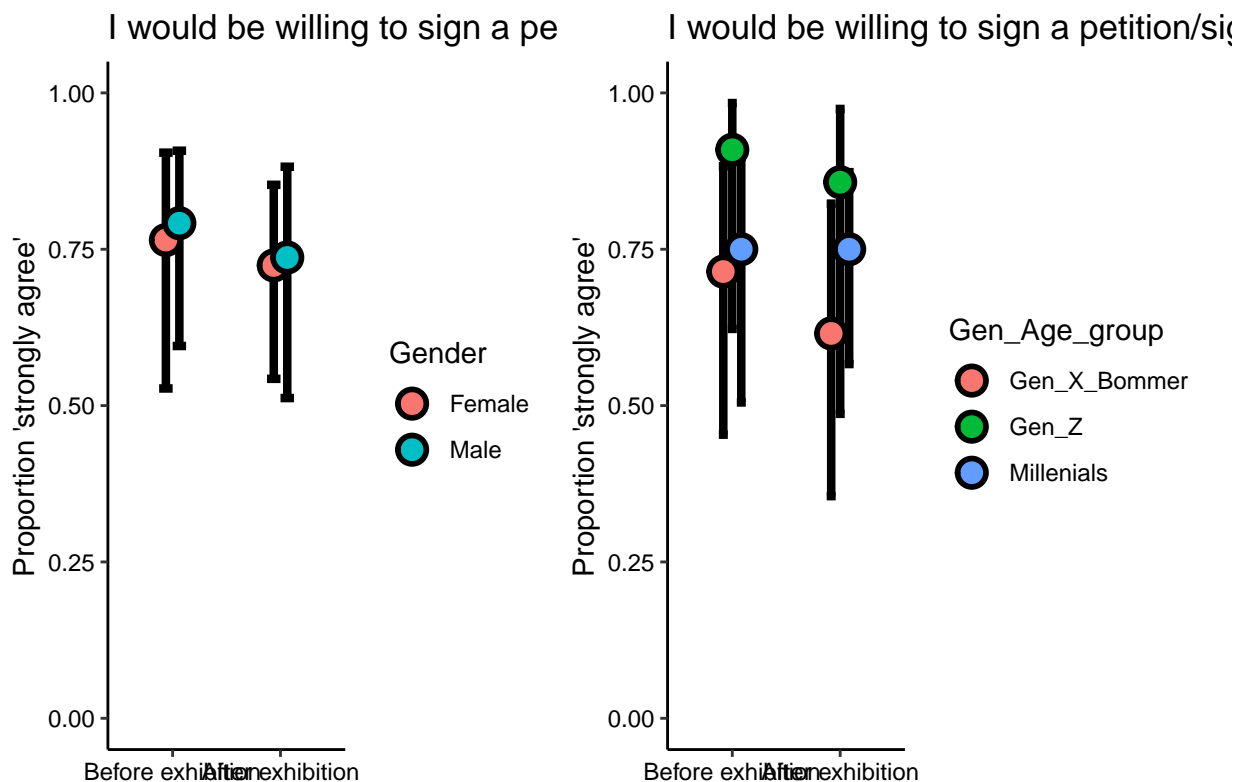


```r
## Plot results (gender and treatment)
plot_data <- Q8b_data %>%
```

```r
  dplyr::group_by(Gender, Treatment, Action) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(value), n = n()), .groups = "drop")

gender <- ggplot() +
  geom_errorbar(data = plot_data, aes(x = Treatment, ymin = lower, ymax = upper, group = Gender),
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data, aes(x = Treatment, y = mean, fill = Gender),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "I would be willing to sign a petition/sign up for newsletter") +
  facet_wrap(facets = ~Action) +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))

## Plot results (gender and treatment)
plot_data_age <- Q8b_data %>%
  dplyr::group_by(Gen_Age_group, Treatment, Action) %>%
  dplyr::summarise(binom::binom.wilson(x = sum(value), n = n()), .groups = "drop")

age <- ggplot() +
  geom_errorbar(data = plot_data_age, aes(x = Treatment, ymin = lower, ymax = upper, group = Gen_Age_gr
                size = 1.5, width = 0.25, position = position_dodge(width = 0.25)) +
  geom_point(data = plot_data_age, aes(x = Treatment, y = mean, fill = Gen_Age_group),
             shape = 21, size = 4, stroke = 1.5, position = position_dodge(width = 0.25)) +
  scale_y_continuous(limits = c(0, 1)) +
  scale_x_discrete(labels = c("Before exhibition", "After exhibition")) +
  labs(y = "Proportion 'strongly agree'", x = "",
       title = "I would be willing to sign a petition/sign up for newsletter") +
  facet_wrap(facets = ~Action) +
  theme_classic() +
  theme(axis.text = element_text(colour = "black"))

library(patchwork)
gender + age
```

Presentation plot

What are the key take-aways we want to present?

Plot 1: Slope graph of responses before/after treatment.

This shows 2 key pieces of information:

1. Males had lower support than females.
2. Males had stronger response to exhibition than females.

```
raw_data <- clean_data %>%
  dplyr::mutate(across(.cols = c(Microplastic_importance_LK, Viable_action_LK, Personal_responsibility_
                                 Political_responsibility_LK, Climate_importance_LK),
                      .fns = ~ dplyr::case_when(. == 7 ~ 1,
                                                . != 7 ~ 0))) %>%
  dplyr::select(ID = X_id, Gender, Treatment, Microplastic_importance_LK, Viable_action_LK, Personal_re
                Political_responsibility_LK, Climate_importance_LK) %>%
  tidyr::pivot_longer(cols = c(Microplastic_importance_LK, Viable_action_LK, Personal_responsibility_LK
                               Political_responsibility_LK, Climate_importance_LK)) %>%
  group_by(ID) %>%
  dplyr::mutate(QuestionNumber = paste0("Q", 1:n()))

plot_data <- raw_data %>%
```

```
    group_by(Gender, Treatment, QuestionNumber) %>%
    summarise(binom::binom.wilson(x = sum(value), n = n()), .groups = "drop") %>%
    dplyr::mutate(Treatment = as.numeric(Treatment),
                  sig = QuestionNumber == "Q3" & Gender == "Male")

text_data <- data.frame(x = c(0.45, 0.45, 0.45),
                        y = c(37, 92, 20),
                        Gender = c("Female", "Male", "Male"),
                        label = c("**<span style = 'color: #8EA4D2; font-weight: bold'>Female</span> res
                                  "**<span style = 'color: #4C9F70; font-weight: bold'>Male</span> respo
                                  "**<span style = 'color: red; font-weight: bold'>Personal responsibili

plot_data %>%
  {ggplot() +
      geom_line(data = ., aes(x = Treatment, y = mean*100, group = QuestionNumber, colour = sig), size =
      geom_point(data = ., aes(x = Treatment, y = mean*100, group = QuestionNumber, colour = sig), size
      geom_text(data = filter(., Treatment == 0), aes(x = Treatment - 0.1,
                                                      y = mean*100,
                                                      label = QuestionNumber,
                                                      colour = sig), size = 4) +

      #Label for female result
      ggtext::geom_richtext(data = text_data,
                            aes(x = x, y = y, label = label),
                            fill = NA, label.color = NA, colour = "black") +
      scale_colour_manual(values = c("grey55", "red")) +
      scale_x_continuous(breaks = c(0, 1),
                         limits = c(-0.25, 1.25),
                         labels = c("Before", "After"), name = "") +
      scale_y_continuous(name = "Percentage 'strongly agree'",
                         limits = c(0, 100),
                         breaks = seq(0, 100, 25),
                         labels = paste0(seq(0, 100, 25), "%")) +
      facet_wrap(facets = ~Gender, ) +
      theme_classic() +
      theme(axis.text = element_text(colour = "black"),
            legend.position = "none")}
```
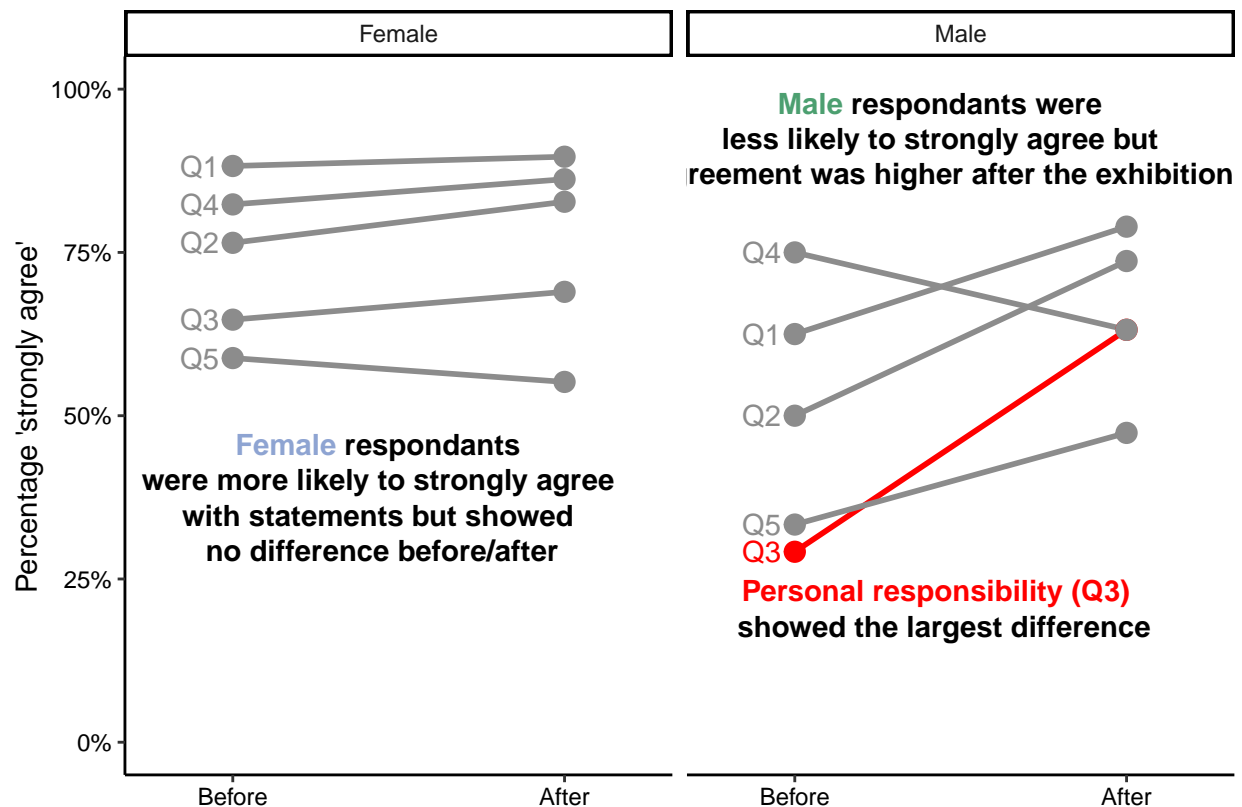
```
ggplot2::ggsave(file = here::here("./plot/Gender_effect.png"), height = 5, width = 10, dpi = 300)
```

Plot 1b: gt table to link 'Q1' to actual question wording

Create data frame of questions

```
Q_df <- data.frame(Qnumber = paste0("Q", 1:5),
                   Wording = c("Microplastic pollution is a major problem that needs to be addressed",
                               "We as humans are able to act against microplastic pollution",
                               "I have a personal responsibility to change my behaviour to avoid micropl
                               "Politicians have a responsibility to implement legislation to avoid mic
                               "Climate change is the most important problem facing our society"))

(Q_df %>%
    gt() %>%
    #Rename columns
    cols_label(Qnumber = "",
               Wording = "") %>%
    #Add a table title
    #Notice the `md` function allows us to write the title using markdown syntax (which allows HTML)
    tab_header(title = md("Survey questions")) %>%
    #Add a data source footnote
    tab_source_note(source_note = "Data: Ocean.Now! 'In Your Face' survey") %>%
    #Apply new style to all column headers
    tab_style(
```

```
    locations = cells_column_labels(columns = everything()),
    style     = list(
      #Give a thick border below
      cell_borders(sides = "bottom", weight = px(3)),
      #Make text bold
      cell_text(weight = "bold")
    )
) %>%
#Apply different style to the title
tab_style(
  locations = cells_title(groups = "title"),
  style     = list(
    cell_text(weight = "bold", size = 24)
  )
))
```

<div align="center">Survey questions</div>

Q1	Microplastic pollution is a major problem that needs to be addressed
Q2	We as humans are able to act against microplastic pollution
Q3	I have a personal responsibility to change my behaviour to avoid microplastic pollution
Q4	Politicians have a responsibility to implement legislation to avoid microplastic pollution
Q5	Climate change is the most important problem facing our society

Data: Ocean.Now! 'In Your Face' survey