psy504_glm

February 12, 2025

- Assignment requirements:
  - If you are using Github (recommended), make sure to commit and push your work to GitHub regularly, at least after each exercise. Write short and informative commit messages, and share the link to your assignment with me. If not, you can also send me the **rmd** & **rendered file** via Canvas.
  - In this assignment, you will not need to code from scratch. Rather, you'll need to fill in code where needed. This assignment has a logisitic regression implementation for a scenario from EDA down to model comparison (and would be useful for whenever you may encounter such a situation in the future).
  - I want the assignments to begin reflecting a bit more of how you'd be doing things on your own, where you have some prior knowledge and you figure other things out (by referring to documentation, etc.) . In addition to the rmd, I also want you to submit to me **notes** of anything new that you learn while finishing the assignment. And any pain-points, and we'll discuss more.
- Note:
  - If you are fitting a model, display the model output in a neatly formatted table. (The `gt` `tidy` and `kable` functions can help!). `Modelsummary` also looks good(https://vincentarelbundock.github.io/modelsummary/articles/modelsummary.html)
  - Make sure that your plots are clearly labeled – for all axes, titles, etc.

0.1 Data: General Social Survey

The General Social Survey (GSS) has been used to measure trends in attitudes and behaviors in American society since 1972. In addition to collecting demographic information, the survey includes questions used to gauge attitudes about government spending priorities, confidence in institutions, lifestyle, and many other topics. A full description of the survey may be found here.

The data for this lab are from the 2016 General Social Survey. The original data set contains 2867 observations and 935 variables. We will use and abbreviated data set that includes the following variables:

natmass: Respondent's answer to the following prompt:

"We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount…are we spending too much, too little, or about the right amount on mass transportation?"

age: Age in years.

sex: Sex recorded as male or female

sei10: Socioeconomic index from 0 to 100

region: Region where interview took place

polviews: Respondent's answer to the following prompt:

"We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal - point 1 - to extremely conservative - point 7. Where would you place yourself on this scale?"

The data are in gss2016.csv in the data folder.

1 EDA

[35]:
```
suppressPackageStartupMessages({
    library(dplyr)
    library(ggplot2)
    library(readr)
    library(tidyr)
    library(knitr)
    library(easystats)
    library(broom)
    library(emmeans)
    library(marginaleffects)
    library(performance)
    library(arm)
    library(modelsummary)
})
```

[3]:
```
data <- read.csv("gss2016.csv")
head(data)
```

A data.frame: 6 × 6

	natmass <chr>	age <chr>	sex <chr>	sei10 <dbl>	region <chr>	polviews <chr>
1	Too little	47	Male	87.9	New england	Moderate
2	Too little	61	Male	38.3	New england	Liberal
3	Too much	43	Female	21.8	New england	Moderate
4	Too little	55	Female	39.7	New england	Slightly liberal
5	About right	53	Female	44.6	New england	Slightly liberal
6	Too little	50	Male	80.7	New england	Slightly liberal

[4]:
```
data <- data %>%
  mutate(mass_trans_spend_right = (natmass == "About right"))
head(data)
```

A data.frame: 6 × 7

	natmass <chr>	age <chr>	sex <chr>	sei10 <dbl>	region <chr>	polviews <chr>	mass_trans_ <lgl>
1	Too little	47	Male	87.9	New england	Moderate	FALSE
2	Too little	61	Male	38.3	New england	Liberal	FALSE
3	Too much	43	Female	21.8	New england	Moderate	FALSE
4	Too little	55	Female	39.7	New england	Slightly liberal	FALSE
5	About right	53	Female	44.6	New england	Slightly liberal	TRUE
6	Too little	50	Male	80.7	New england	Slightly liberal	FALSE

[5]:

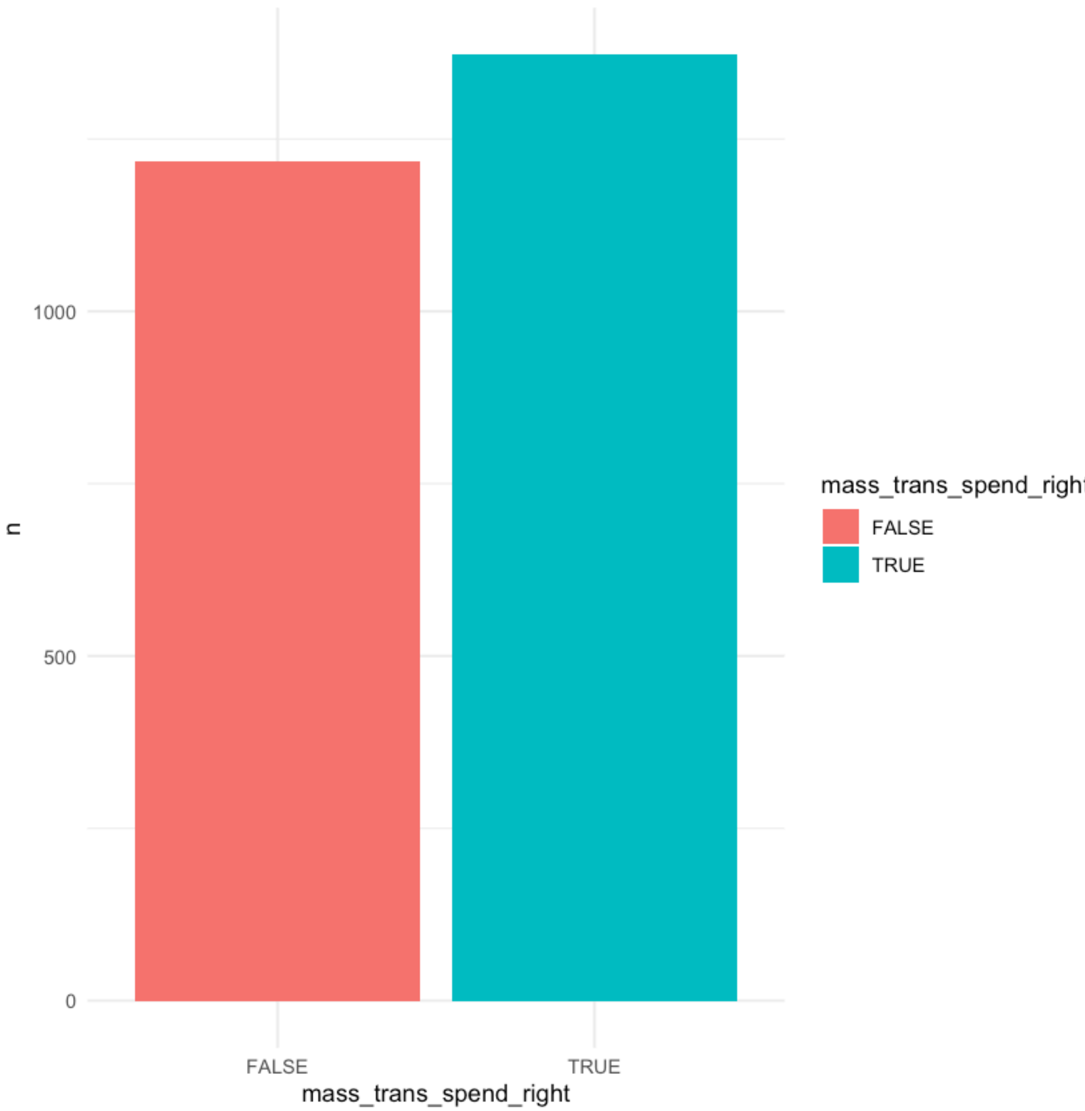
```
#Get proportions
mass_spend_summary <- data %>%
  count(mass_trans_spend_right) %>%
  mutate(proportion = n / sum(n))
mass_spend_summary
```

A data.frame: 2 × 3

mass_trans_spend_right <lgl>	n <int>	proportion <dbl>
FALSE	1218	0.4702703
TRUE	1372	0.5297297

[6]:

```
data %>%
  count(mass_trans_spend_right) %>%  # Count occurrences of TRUE/FALSE
  ggplot(aes(x = mass_trans_spend_right, y = n, fill = mass_trans_spend_right))␣
↪+
  geom_col() +
  theme_minimal()
```



[8]:
```
data %>%
  count(polviews)
```

A data.frame: 7 × 2

polviews	n
<chr>	<int>
Conservative	395
Extremely liberal	133
Extrmly conservative	104
Liberal	336
Moderate	974
Slghtly conservative	358
Slightly liberal	290

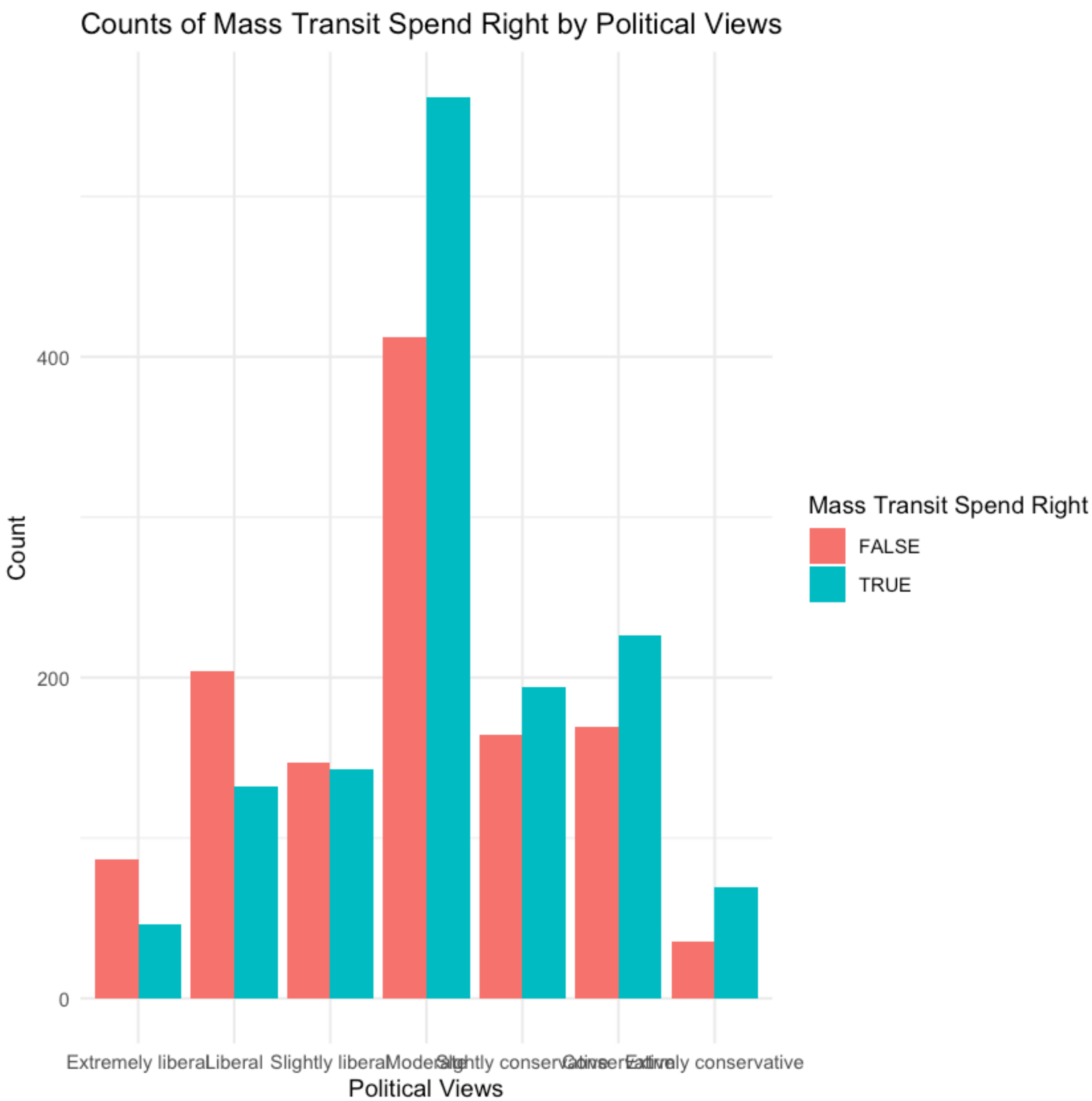
the most freq is 'Moderate'

[9]:
```r
data <- data %>%
  mutate(polviews = factor(polviews,
                           levels = c("Extremely liberal", "Liberal", "Slightly␣
↪liberal", "Moderate",
                                      "Slghtly conservative", "Conservative",␣
↪"Extrmly conservative"),
                           ordered = TRUE))

head(data)
```

A data.frame: 6 × 7

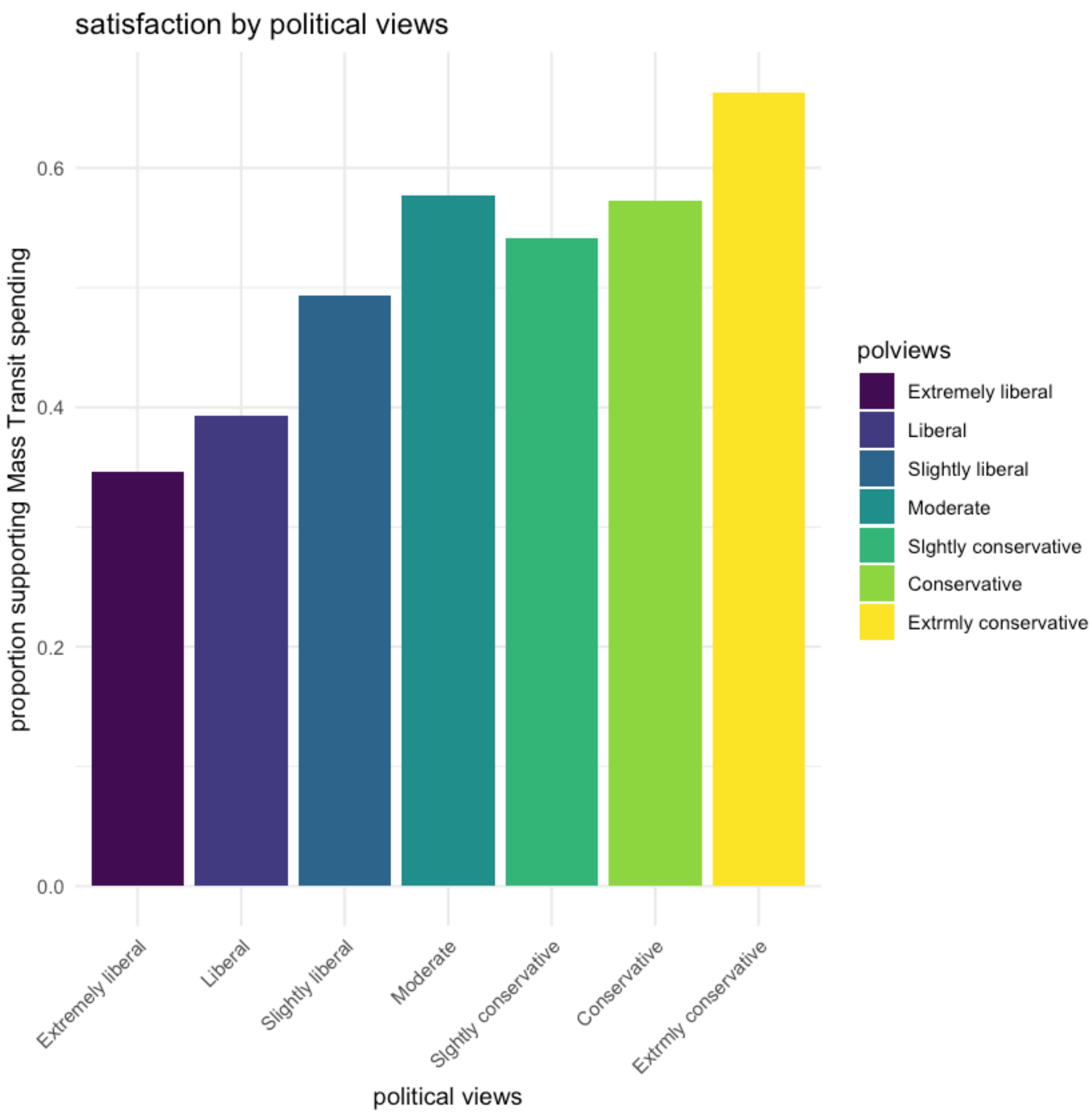
	natmass <chr>	age <chr>	sex <chr>	sei10 <dbl>	region <chr>	polviews <ord>	mass_trans_ <lgl>
1	Too little	47	Male	87.9	New england	Moderate	FALSE
2	Too little	61	Male	38.3	New england	Liberal	FALSE
3	Too much	43	Female	21.8	New england	Moderate	FALSE
4	Too little	55	Female	39.7	New england	Slightly liberal	FALSE
5	About right	53	Female	44.6	New england	Slightly liberal	TRUE
6	Too little	50	Male	80.7	New england	Slightly liberal	FALSE

[10]:
```r
data %>%
  count(polviews, mass_trans_spend_right) %>%  # Count occurrences within␣
↪polviews
  ggplot(aes(x = polviews, y = n, fill = as.factor(mass_trans_spend_right))) +
  geom_col(position = "dodge") +
  labs(x = "Political Views", y = "Count", fill = "Mass Transit Spend Right",
       title = "Counts of Mass Transit Spend Right by Political Views") +
  theme_minimal()
```



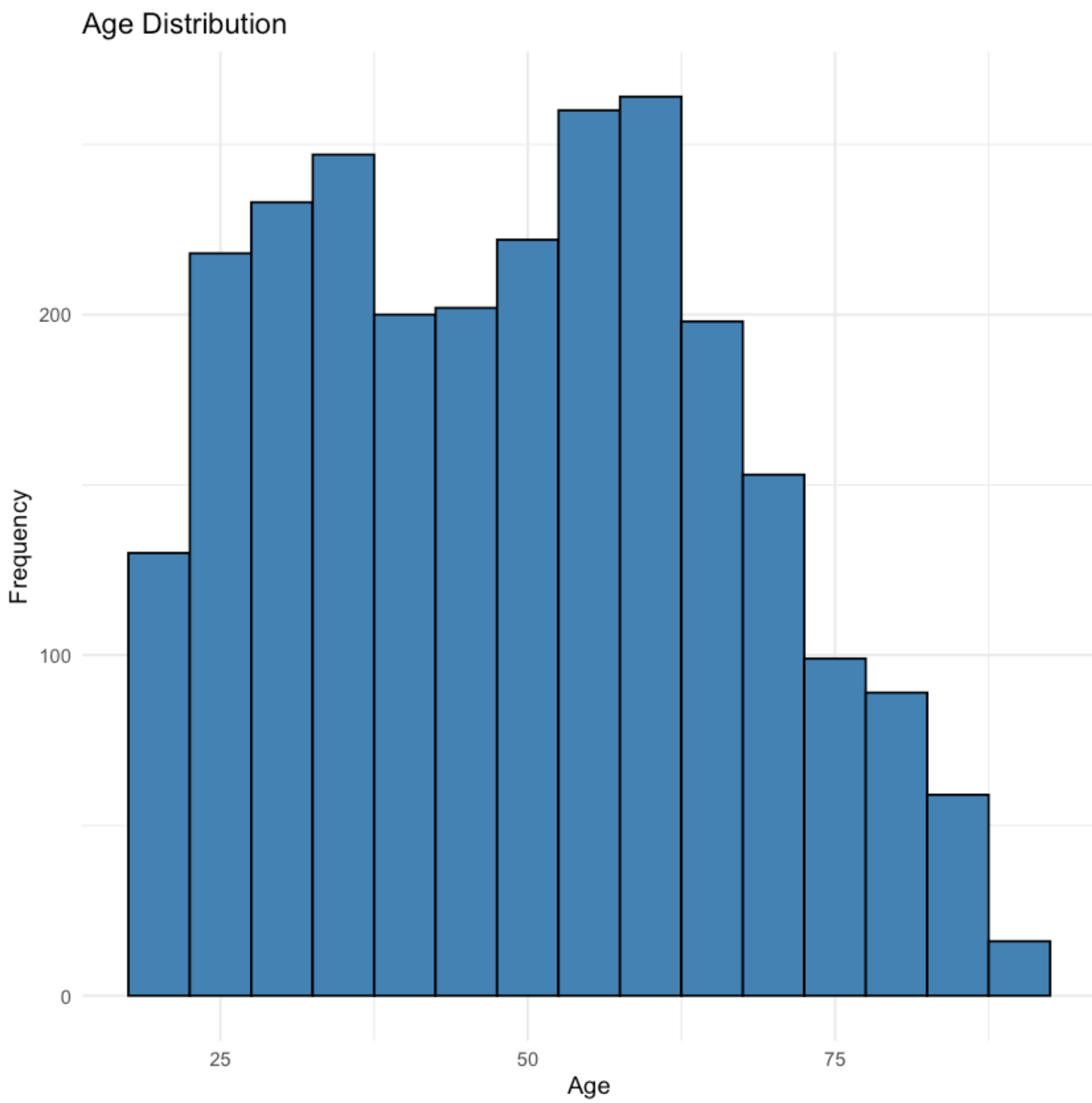
satisfaction by political views

[11]:
```
data %>%
  group_by(polviews) %>%
  summarise(prop_true = mean(mass_trans_spend_right)) %>%  # Calculate proportion of TRUE
  ggplot(aes(x = polviews, y = prop_true, fill = polviews)) +
  geom_col() +
  labs(x = "political views", y = "proportion supporting Mass Transit spending",
       title = "satisfaction by political views") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



[12]:
```
data <- data %>%
  mutate(age = ifelse(age == "89 or older", "89", age),
         age = as.numeric(age))
```

[13]:
```
# freq dist of age
ggplot(data, aes(x = age)) +
  geom_histogram(binwidth = 5, fill = "steelblue", color = "black") +  # Adjust␣
 ↪binwidth as needed
  labs(title = "Age Distribution", x = "Age", y = "Frequency") +
  theme_minimal()
```



2 Logistic regression

[14]:
```
glm_intercept <- glm(mass_trans_spend_right ~ 1, data = data, family = binomial)

summary(glm_intercept)
```

```
Call:
glm(formula = mass_trans_spend_right ~ 1, family = binomial,
    data = data)

Coefficients:
```

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.11906    0.03937   3.024  0.00249 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3581.3  on 2589  degrees of freedom
Residual deviance: 3581.3  on 2589  degrees of freedom
AIC: 3583.3

Number of Fisher Scoring iterations: 3
```

[15]:
```
b0 = coef(glm_intercept)
b0_transformed <- exp(b0) / (1 + exp(b0)) # logistic transform
ci_lower = b0 - 1.96 * 0.0393685
ci_upper = b0 + 1.96 * 0.0393685

#transforming confidence intervals of coefficients into probabilities
p_lower = exp(ci_lower) / (1 + exp(ci_lower))
p_upper = exp(ci_upper) / (1 + exp(ci_upper))

print(b0_transformed)
print(p_lower)
print(p_upper)
```

```
(Intercept)
  0.5297297
(Intercept)
  0.5104727
(Intercept)
  0.5488986
```

‘b0_transformed’ shows that the probability of mass_trans_spend_right is TRUE is around 53%, and is significantly different from 50%, suggesting people in general are satisfied with the spending. This is also evident from CIs which do not include 50%.

[16]:
```
# ensure sex is categorical
data <- data %>%
  mutate(sex = as.factor(sex))

glm_1 <- glm(mass_trans_spend_right ~ age + sex + sei10,
                  data = data,
                  family = binomial)

glm_1 %>%
  tidy() %>%
```

```
  kable()

glm_1 %>%
  tidy(exponentiate = TRUE) %>%
  kable()
```

```
|term        |   estimate| std.error| statistic|   p.value|
|:-----------|----------:|---------:|---------:|---------:|
|(Intercept) |  0.8254509| 0.1395587|  5.914722| 0.0000000|
|age         | -0.0061659| 0.0022824| -2.701502| 0.0069027|
|sexMale     | -0.2557439| 0.0798020| -3.204732| 0.0013519|
|sei10       | -0.0062271| 0.0016609| -3.749229| 0.0001774|
```

```
|term        |  estimate| std.error| statistic|   p.value|
|:-----------|---------:|---------:|---------:|---------:|
|(Intercept) | 2.2829100| 0.1395587|  5.914722| 0.0000000|
|age         | 0.9938530| 0.0022824| -2.701502| 0.0069027|
|sexMale     | 0.7743403| 0.0798020| -3.204732| 0.0013519|
|sei10       | 0.9937922| 0.0016609| -3.749229| 0.0001774|
```

3 effect of Sex

[17]:
```
bsex <- coef(glm_1)["sexMale"]
ci_lower_lo = bsex - 1.96 * 0.0798020
ci_upper_lo = bsex + 1.96 * 0.0798020

ci_lower_or = 1.29 - 1.96 * 0.0798020
ci_upper_or = 1.29 + 1.96 * 0.0798020
emm_sex <- emmeans(glm_1, "sex", type = "response")

print(emm_sex)
```

```
 sex     prob     SE  df asymp.LCL asymp.UCL
 Female 0.559 0.0133 Inf     0.533     0.585
 Male   0.495 0.0147 Inf     0.467     0.524

Confidence level used: 0.95
Intervals are back-transformed from the logit scale
```

4 effect of age

[18]:
```
bage <- coef(glm_1)["age"]

ci_lower_lo_age = bage - 1.96 * 0.0022824
ci_upper_lo_age = bage + 1.96 * 0.0022824

c(LO = bage, CI_Lower = ci_lower_lo_age, CI_Upper = ci_upper_lo_age)

or_age <- exp(bage)
ci_lower_or_age = exp(ci_lower_lo_age)
ci_upper_or_age = exp(ci_upper_lo_age)

c(OR = or_age, CI_Lower = ci_lower_or_age, CI_Upper = ci_upper_or_age)

emm_age <- emmeans(glm_1, "age", at = list(age = seq(min(data$age),
 ↪max(data$age), by = 10)), type = "response")
print(emm_age)
```

LO.age -0.0061659395592894 **CI_Lower.age** -0.0106394435592894 **CI_Upper.age** -0.0016924355592894

OR.age 0.993853030835906 **CI_Lower.age** 0.989416955126246 **CI_Upper.age** 0.998308995802162

```
 age  prob      SE  df asymp.LCL asymp.UCL
  18 0.574 0.01980 Inf     0.535     0.613
  28 0.559 0.01530 Inf     0.529     0.589
  38 0.544 0.01160 Inf     0.521     0.567
  48 0.529 0.00993 Inf     0.509     0.548
  58 0.513 0.01120 Inf     0.491     0.535
  68 0.498 0.01470 Inf     0.469     0.527
  78 0.482 0.01930 Inf     0.445     0.520
  88 0.467 0.02430 Inf     0.420     0.515

Results are averaged over the levels of: sex
Confidence level used: 0.95
Intervals are back-transformed from the logit scale
```

A one unit increase in age is associated with a decrease in the log-odds of being satisfied with spending on mass transportation by -0.0061659 units (95% CI [-0.010, -0.002]), holding all other variables constant. The odds ratio is 0.9938530 (95% CI [0.989, 0.998]) which confirms the negative relationship implied by the log-odds coefficient.

Specifically, for each additional unit of age, the odds of being satisfied with mass transportation spending decrease by a factor of about 0.9938, or approximately 0.62% per unit increase in age, holding other factors constant.

5 effect of SES index

[19]:
```
bses <- coef(glm_1)["sei10"]
ci_lower_lo_ses = bses - 1.96 * 0.0001774
ci_upper_lo_ses = bses + 1.96 * 0.0001774

c(LO = bses, CI_Lower = ci_lower_lo_ses, CI_Upper = ci_upper_lo_ses)

or_ses <- exp(bses)
ci_lower_or_ses = exp(ci_lower_lo_ses)
ci_upper_or_ses = exp(ci_upper_lo_ses)

c(OR = or_ses, CI_Lower = ci_lower_or_ses, CI_Upper = ci_upper_or_ses)

emm_ses <- emmeans(glm_1, "sei10", type = "response")
print(emm_ses)
```

LO.sei10 -0.00622714075833176 **CI_Lower.sei10** -0.00657484475833176 **CI_Upper.sei10** -0.00587943675833176

OR.sei10 0.99379220769999 **CI_Lower.sei10** 0.993446722241023 **CI_Upper.sei10** 0.99413781330652

```
 sei10  prob      SE  df asymp.LCL asymp.UCL
  46.1 0.527 0.00992 Inf     0.508     0.547

Results are averaged over the levels of: sex
Confidence level used: 0.95
Intervals are back-transformed from the logit scale
```

A one unit increase in ses is associated with a decrease in the log-odds of being satisfied with spending on mass transportation by -0.006227 units (95% CI [-0.00657, -0.005879]), holding all other variables constant. The odds ratio is 0.99379 (95% CI [0.993, 0.994]) which confirms the negative relationship implied by the log-odds coefficient.

Specifically, for each additional unit of ses, the odds of being satisfied with mass transportation spending decrease by a factor of about 0.99379, or approximately 0.63% per unit increase in ses, holding other factors constant.

6 marginal effects

Let's examine the results on the probability scale.

[20]:
```
avg_comparisons(glm_1, comparison = "difference") %>%
  kable()
```

```
|term  |contrast      |    estimate| std.error| statistic|    p.value|    s.value|
  conf.low|  conf.high|
```

```
|:-----|:-------------|----------:|---------:|---------:|---------:|---------:
↪|----------:|----------:|
|age   |+1            | -0.0015153| 0.0005579| -2.716128| 0.0066050|  7.242217|␣
↪-0.0026088| -0.0004219|
|sei10 |+1            | -0.0015304| 0.0004039| -3.789362| 0.0001510| 12.692830|␣
↪-0.0023219| -0.0007388|
|sex   |Male - Female | -0.0630688| 0.0196461| -3.210251| 0.0013262|  9.558494|␣
↪-0.1015743| -0.0245632|
```

The marginal effect of age is -0.0015153 (95% CI [-0.0026088, -0.0004219]). So, for each additional unit increase of age, the probability of being satisfied with mass transportation spending decreases by approximately 0.1515 percentage points, holding other factors constant (p = 0.0066050).

The marginal effect of SES is -0.0015304 (95% CI [-0.0023219, -0.0007388]). For each one-unit increase in the socioeconomic index, the probability of being satisfied with mass transportation spending decreases by approximately 0.15304 percentage points, holding other variables constant (p = 0.0001510).

The marginal effect for being male compared to female is -0.0630688 (95% CI [-0.1015743, -0.0245632]). This indicates that males are, on average, about 6.3% percentage points less likely than females to be satisfied with mass transportation spending, holding other factors constant (p = 0.0013262).

7 model comparison

[21]:
```r
fit_2 = glm(mass_trans_spend_right ~ age + sex + sei10,
                 data = data,
                 family = binomial)
fit_3 = glm(mass_trans_spend_right ~ age + sex + sei10 + polviews,
                 data = data,
                 family = binomial)

test_likelihoodratio(fit_2, fit_3) %>% kable()

# equivalently
anova(fit_2, fit_3, test = "Chisq")
```

```
|      |Name  |Model | df| df_diff|     Chi2|  p|
|:-----|:-----|:-----|--:|-------:|--------:|--:|
|fit_2 |fit_2 |glm   |  4|      NA|       NA| NA|
|fit_3 |fit_3 |glm   | 10|       6| 63.02844|  0|
```

A anova: 2 × 5

	Resid. Df <dbl>	Resid. Dev <dbl>	Df <dbl>	Deviance <dbl>	Pr(>Chi) <dbl>
1	2586	3545.796	NA	NA	NA
2	2580	3482.768	6	63.02844	1.089119e-11

the model with 'polviews' is significantly better

8 visualization

[22]:
```
library(ggeffects)

str(data)

data <- data %>%
  mutate(polviews = as.factor(polviews))

data <- data %>%
  mutate(mass_trans_spend_right = as.numeric(mass_trans_spend_right))

fit_3 <- glm(mass_trans_spend_right ~ age + sex + sei10 + polviews,
             data = data,
             family = binomial)
```

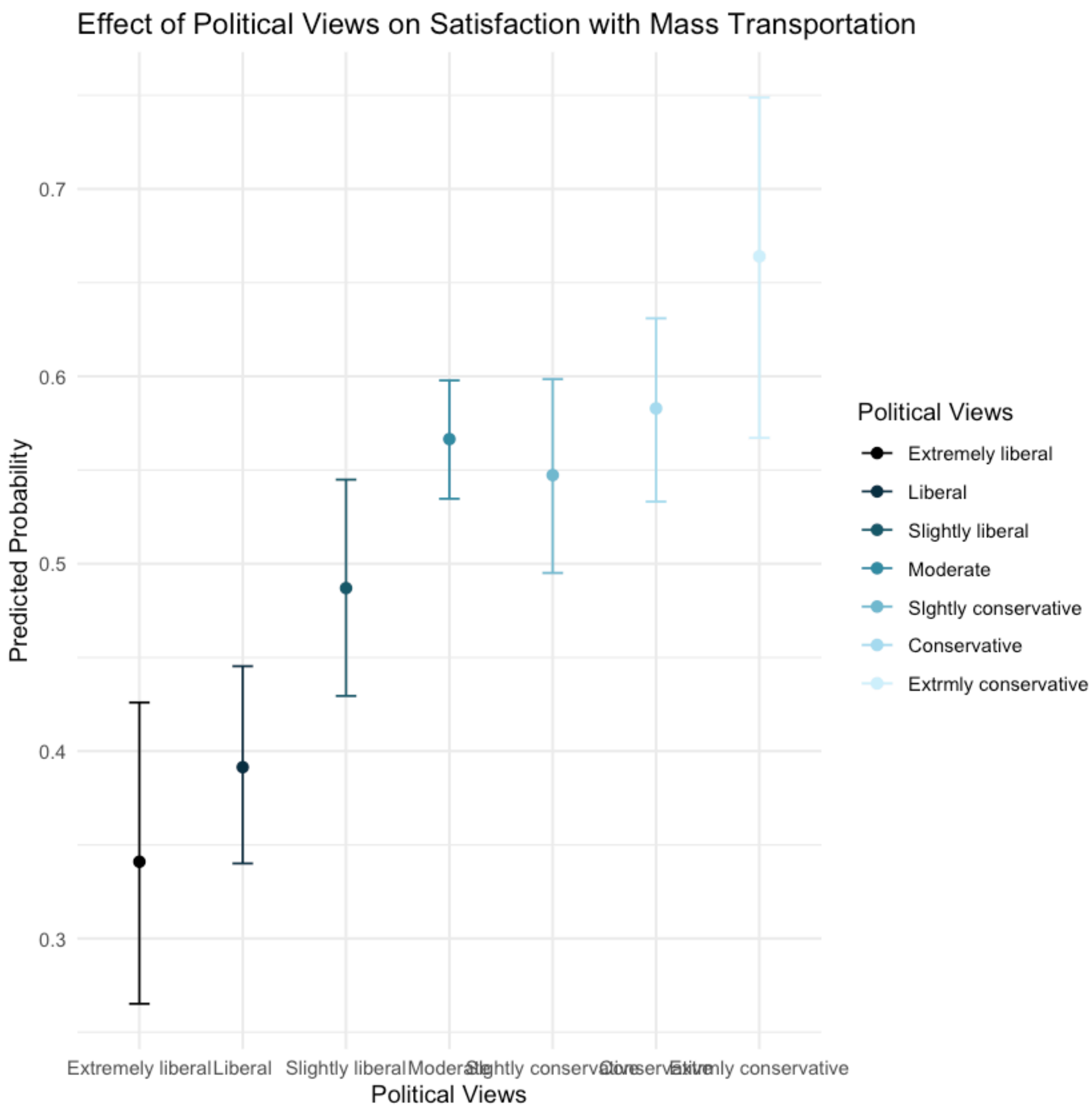
```
Attaching package: ‘ggeffects’


The following object is masked from ‘package:easystats’:

    install_latest



'data.frame':   2590 obs. of  7 variables:
 $ natmass               : chr  "Too little" "Too little" "Too much" "Too
little" …
 $ age                   : num  47 61 43 55 53 50 23 71 86 32 …
 $ sex                   : Factor w/ 2 levels "Female","Male": 2 2 1 1 1 2 1 2 1
2 …
 $ sei10                 : num  87.9 38.3 21.8 39.7 44.6 80.7 20.1 32 13.2 20.8
…
 $ region                : chr  "New england" "New england" "New england" "New
england" …
 $ polviews              : Ord.factor w/ 7 levels "Extremely liberal"<..: 4 2 4
3 3 3 5 6 5 3 …
 $ mass_trans_spend_right: logi  FALSE FALSE FALSE FALSE TRUE FALSE …
```

[23]:
```
colors <- c("Extremely liberal" = "black",
            "Liberal" = "#0e2f44",  # Dark blue
            "Slightly liberal" = "#1d5a6c",  # Less dark blue
            "Moderate" = "#358ca3",  # Medium blue
            "Slghtly conservative" = "#71b9d1",  # Light blue
            "Conservative" = "#a6dcef",  # Lighter blue
            "Extrmly conservative" = "#d0f0fd")  # Very light blue
```

```
pp_pol <- ggemmeans(fit_3, terms = "polviews")
print(pp_pol)
```

```
# Predicted probabilities of mass_trans_spend_right

polviews              | Predicted |     95% CI
----------------------------------------------
Extremely liberal     |      0.34 | 0.27, 0.43
Liberal               |      0.39 | 0.34, 0.45
Slightly liberal      |      0.49 | 0.43, 0.54
Moderate              |      0.57 | 0.53, 0.60
Slghtly conservative  |      0.55 | 0.50, 0.60
Conservative          |      0.58 | 0.53, 0.63
Extrmly conservative  |      0.66 | 0.57, 0.75


Adjusted for:

*   age = 48.90

* sei10 = 46.07
```

[24]:
```
# Adjusted plot with gradient colors
pol_plot <- ggplot(pp_pol, aes(x = x, y = predicted, color = x)) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = 0.2) +
  scale_color_manual(values = colors) +
  labs(title = "Effect of Political Views on Satisfaction with Mass␣
↪Transportation",
       x = "Political Views", y = "Predicted Probability",
       color = "Political Views") +
  theme_minimal()

pol_plot
```

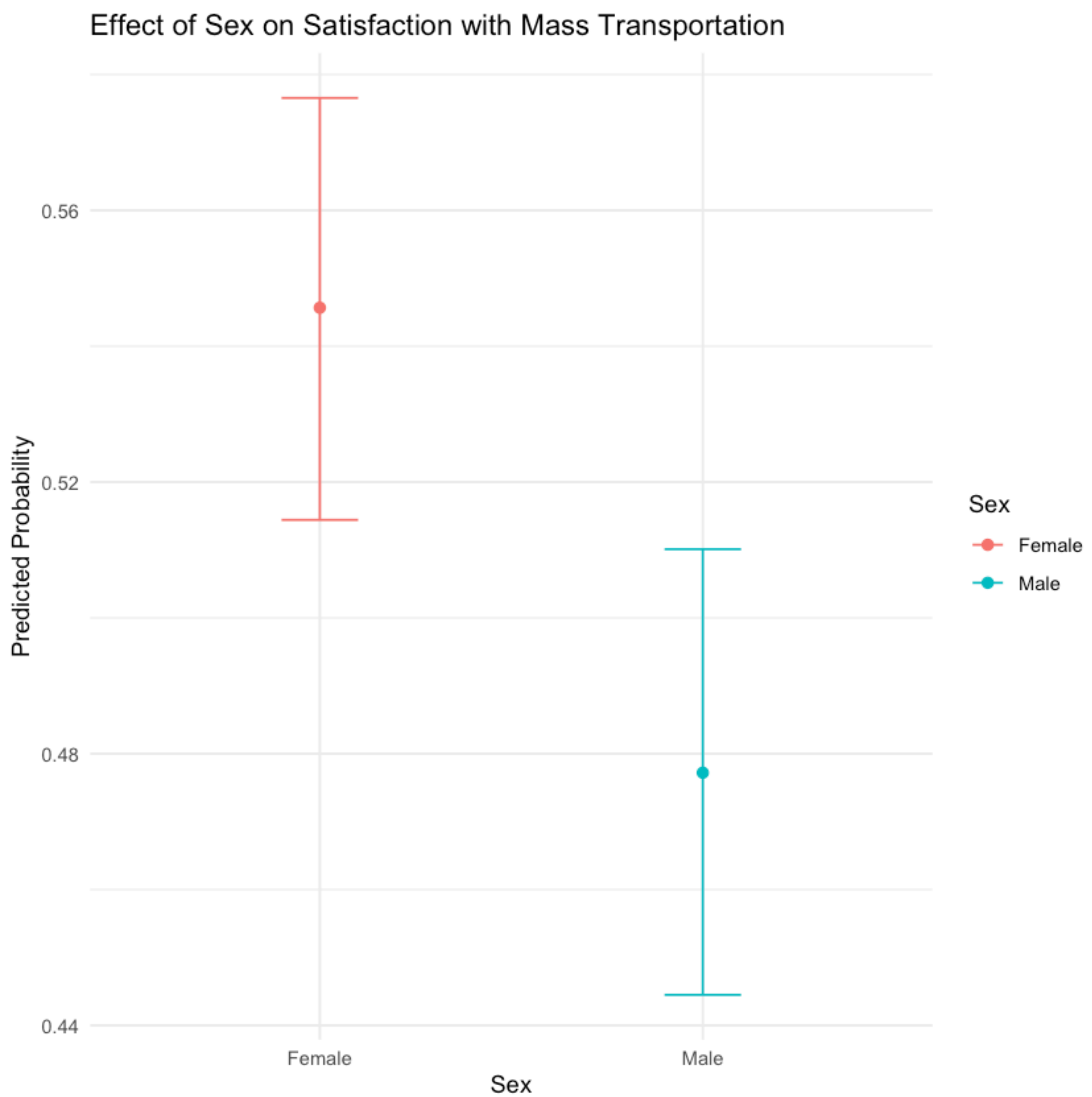


The above plot shows the probability of being satisfied with Mass transportation increases when political views become more conservative, when controlling for age and SES.

[25]:
```
pp_sex <- ggemmeans(fit_3, terms = c("sex"))

sex_plot <- ggplot(pp_sex, aes(x = x, y = predicted, color = x)) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = 0.2) +
  labs(title = "Effect of Sex on Satisfaction with Mass Transportation",
       x = "Sex", y = "Predicted Probability",
       color = "Sex") +
  theme_minimal()
```

```r
print(pp_sex)
sex_plot
```

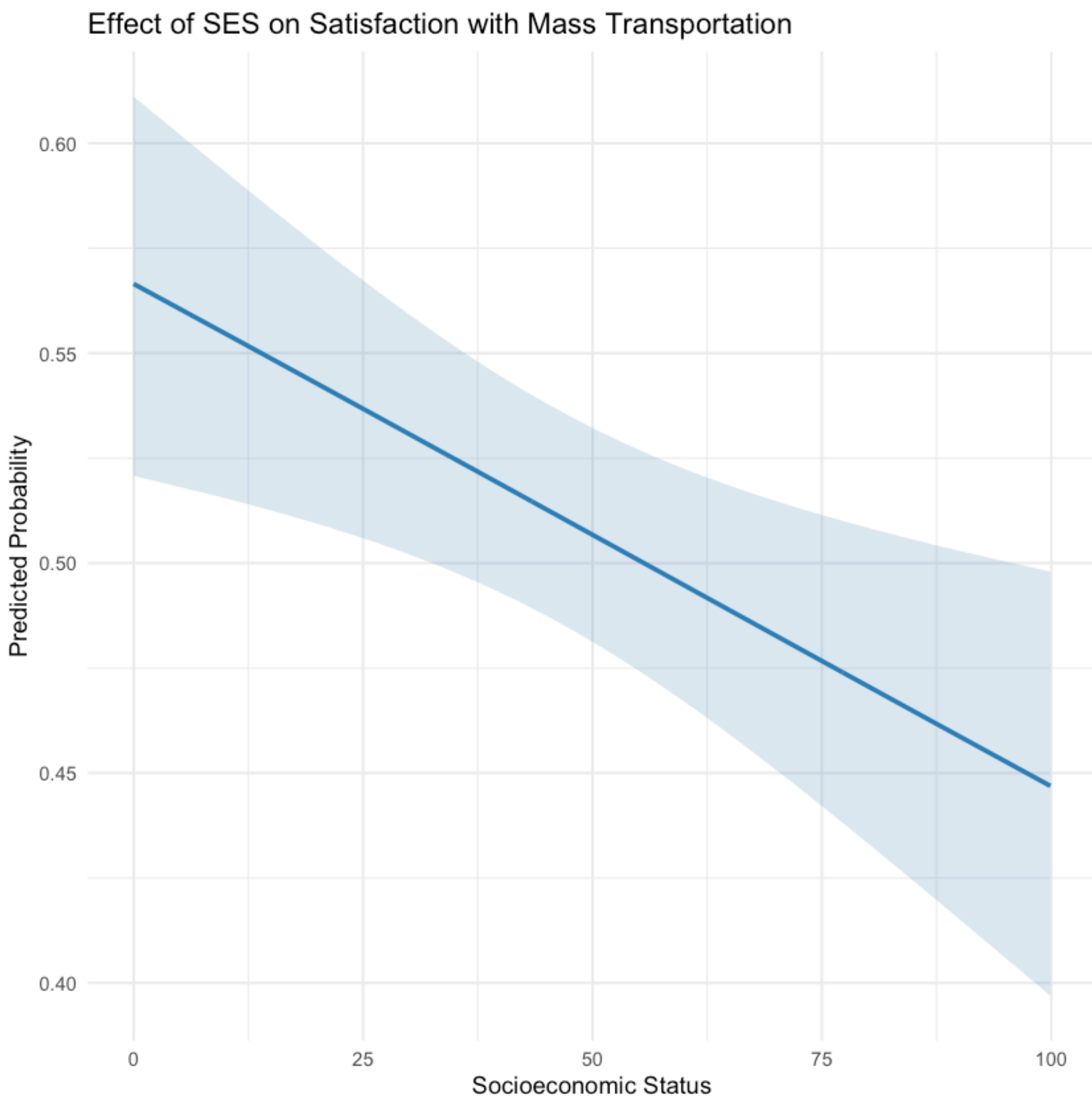
```
# Predicted probabilities of mass_trans_spend_right

sex    | Predicted |     95% CI
-------------------------------
Female |      0.55 | 0.51, 0.58
Male   |      0.48 | 0.44, 0.51


Adjusted for:
*   age = 48.90
* sei10 = 46.07
```



The above plot shows that women are in general more satisfied with Mass transportation than man, when controlling for age and SES.

[27]:
```
pp_ses <- ggemmeans(fit_3, terms = "sei10 [all]")

ses_plot <-  ggplot(pp_ses, aes(x = x, y = predicted)) +
  geom_line(color = "#2c7fb8", size = 1) +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high), fill = "#2c7fb8", alpha =␣
↪0.2) +  # Add a confidence interval band
  labs(title = "Effect of SES on Satisfaction with Mass Transportation",
       x = "Socioeconomic Status", y = "Predicted Probability") +
  theme_minimal() +
  theme(legend.position = "none")

print(pp_ses)
ses_plot
```

```
# Predicted probabilities of mass_trans_spend_right

sei10 | Predicted |     95% CI
------------------------------
 0.00 |      0.57 | 0.52, 0.61
24.10 |      0.54 | 0.51, 0.57
31.60 |      0.53 | 0.50, 0.56
39.50 |      0.52 | 0.49, 0.55
50.40 |      0.51 | 0.48, 0.53
59.40 |      0.50 | 0.47, 0.52
70.30 |      0.48 | 0.45, 0.51
99.90 |      0.45 | 0.40, 0.50


Adjusted for:

* age = 48.90


Not all rows are shown in the output. Use `print(…, n = Inf)` to show
  all rows.
```



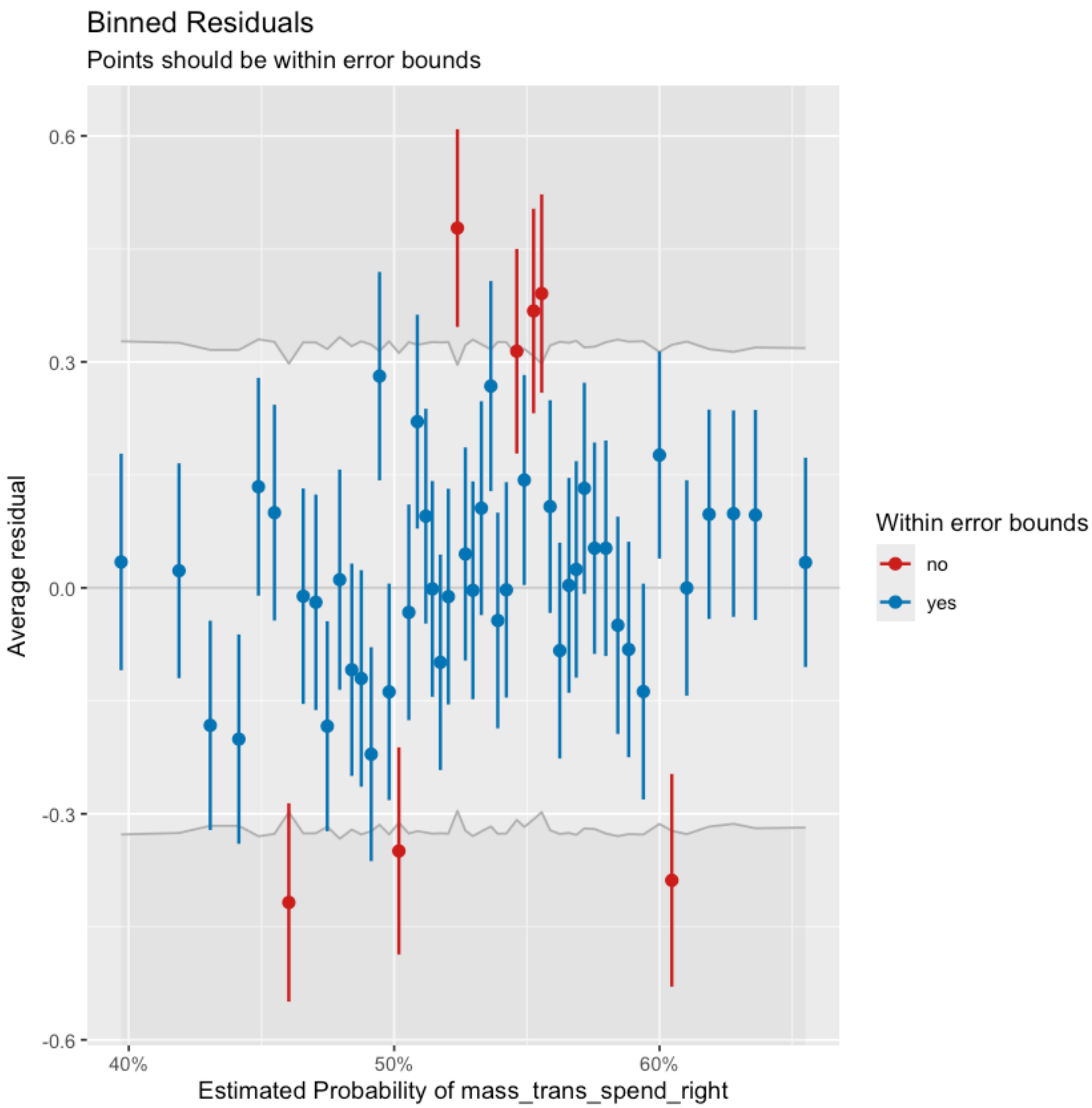
The above plot shows that the satisfaction for Mass transportation decreases with increase in SES, while controlling for age.

9 model assumptions

Is the logistic model a good choice for this data?

[34]:
```
bin_res = binned_residuals(fit_2)
print(bin_res)
plot(bin_res)
```

```
Warning: About 86% of the residuals are inside the error bounds (~95% or
higher would be good).
```



The model does not seem to fit the data well, because of low proportion of data (86% as opposed to the ideal >95%) within the error bound.

[29]:
```
r2_mcfadden(fit_2)
```

```
# R2 for Generalized Linear Regression
       R2: 0.010
  adj. R2: 0.009
```

The model explains 1% of variance, suggesting a poor fit.

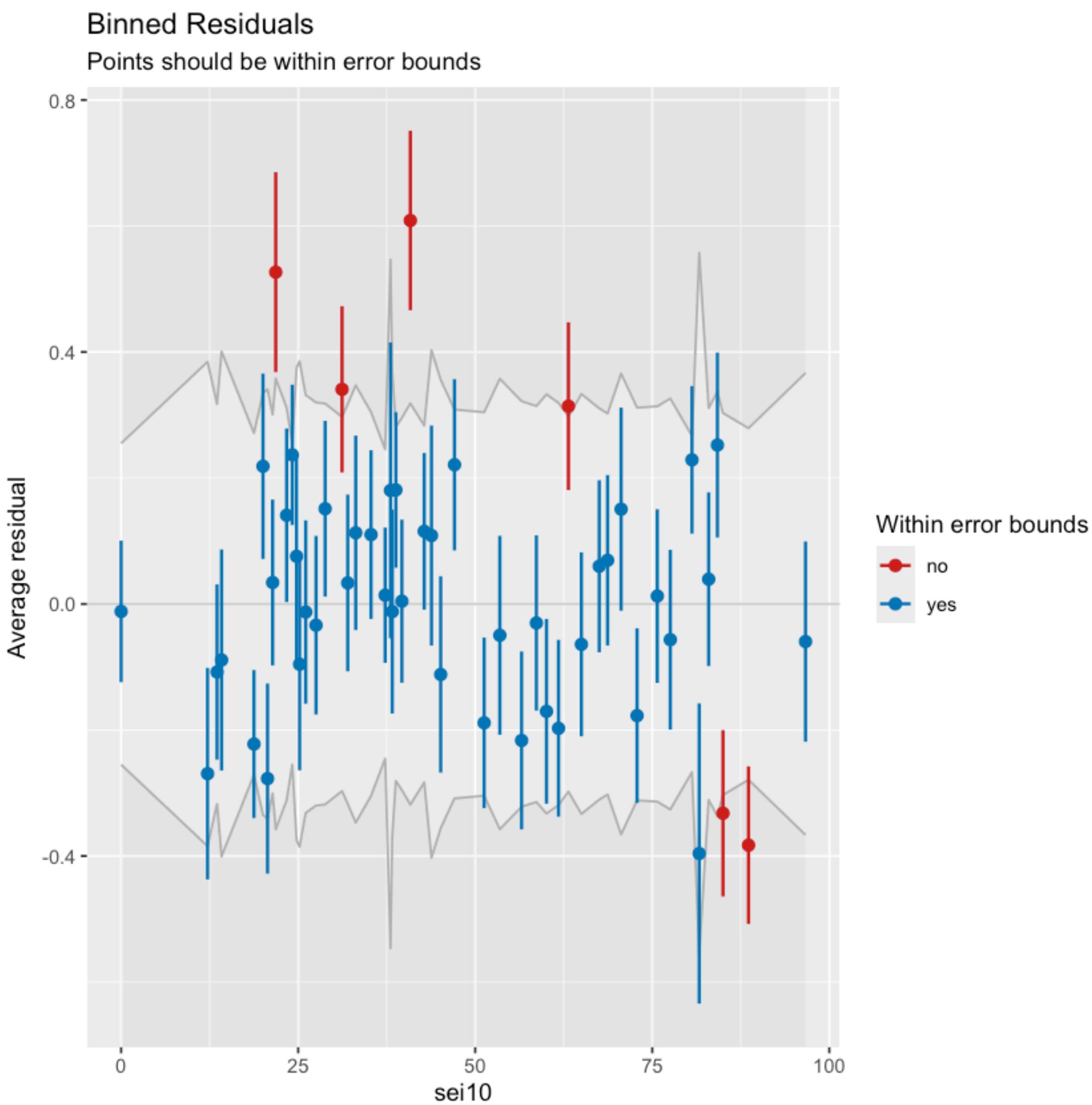
[30]:
```
binned_residuals(fit_2, term="sei10")
binned_residuals(fit_2, term="age")

binned_residuals(fit_2, term="sei10") %>% plot(show_dots=TRUE)
binned_residuals(fit_2, term="age") %>% plot(show_dots=TRUE)
```

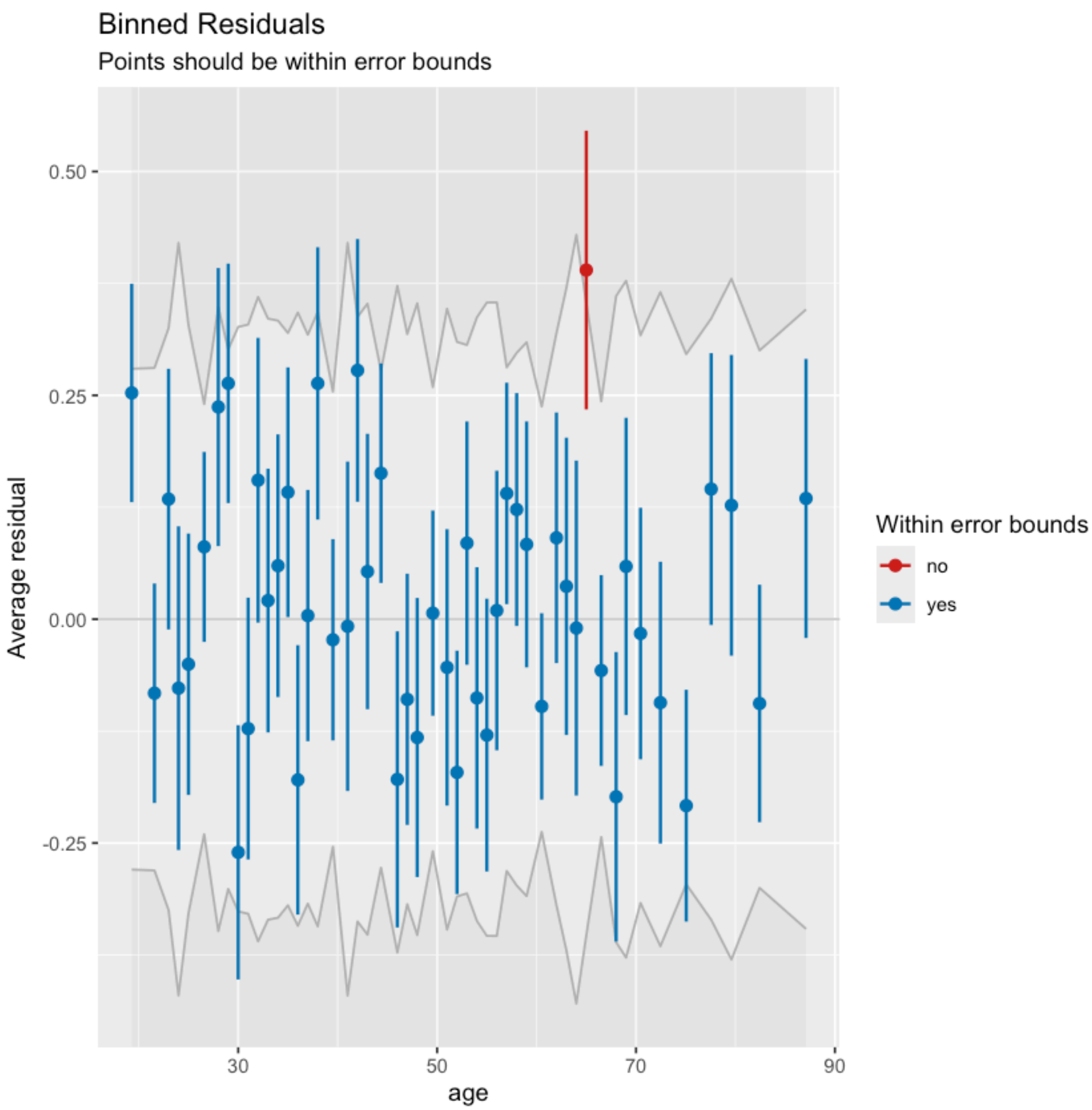
A binned_residuals: 51 × 9

	xbar	ybar	n	x.lo	x.hi	se	CI_l
	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl
conf_int	0.00000	-0.011793983	80	0.0	0.0	0.2550301	-0.12
conf_int1	12.20541	-0.269309119	37	9.0	12.6	0.3845044	-0.43
conf_int2	13.54259	-0.107946954	54	13.2	13.6	0.3174032	-0.24
conf_int3	14.18235	-0.088890080	34	14.0	14.8	0.4006327	-0.26
conf_int4	18.75600	-0.222304869	75	15.8	19.6	0.2713304	-0.33
conf_int5	20.03182	0.218663094	44	19.7	20.1	0.3347770	0.071
conf_int6	20.67333	-0.277131083	45	20.5	20.7	0.3409253	-0.42
conf_int7	21.39322	0.034109141	59	20.8	21.6	0.3007066	-0.09
conf_int8	21.83103	0.526726461	29	21.8	22.4	0.3579303	0.368
conf_int9	23.36981	0.140830987	53	22.9	23.8	0.3116481	0.003
conf_int10	24.16842	0.236713173	76	24.0	24.2	0.2549783	0.125
conf_int11	24.77368	0.075699992	38	24.6	25.0	0.3750762	-0.08
conf_int12	25.20270	-0.095667272	37	25.1	25.4	0.3852950	-0.26
conf_int13	26.06531	-0.012777085	49	25.7	26.8	0.3316817	-0.15
conf_int14	27.52115	-0.033649173	52	26.9	28.4	0.3199773	-0.17
conf_int15	28.80769	0.151224428	52	28.5	29.6	0.3181639	0.011
conf_int16	31.19630	0.340733568	54	29.9	31.6	0.2970537	0.208
conf_int17	31.99434	0.033395797	53	31.9	32.0	0.3172123	-0.10
conf_int18	33.12558	0.113020911	43	32.6	34.4	0.3471490	-0.04
conf_int19	35.29123	0.110167765	57	34.5	35.8	0.3046176	-0.02
conf_int20	37.28315	0.013944083	89	36.2	37.6	0.2456516	-0.09
conf_int21	38.03889	0.180317687	18	37.8	38.1	0.5465760	-0.05
conf_int22	38.28750	-0.012005609	40	38.2	38.3	0.3688487	-0.17
conf_int23	38.79385	0.181048581	65	38.4	38.8	0.2809402	0.057
conf_int24	39.64516	0.004424233	62	39.0	39.7	0.2954619	-0.12
conf_int25	40.83714	0.608763998	35	39.9	41.1	0.3185876	0.466
conf_int26	42.80000	0.115359593	66	41.4	43.0	0.2832513	-0.00
conf_int27	43.82353	0.108690330	34	43.4	43.9	0.4030259	-0.06
conf_int28	45.12558	-0.111853001	43	44.5	45.4	0.3559469	-0.26
conf_int29	47.07222	0.221015977	54	46.0	49.4	0.3086951	0.085
conf_int30	51.26071	-0.188652521	56	49.6	52.0	0.3043499	-0.32
conf_int31	53.47381	-0.049632650	42	52.1	54.6	0.3575811	-0.20
conf_int32	56.54902	-0.216585677	51	55.1	57.5	0.3218849	-0.35
conf_int33	58.63704	-0.030175847	54	57.8	59.1	0.3142450	-0.16
conf_int34	60.06458	-0.170403945	48	59.4	60.5	0.3328321	-0.31
conf_int35	61.76346	-0.197321220	52	61.2	62.4	0.3191634	-0.33
conf_int36	63.16182	0.314008214	55	62.6	63.7	0.2977596	0.180
conf_int37	64.99388	-0.064046370	49	64.2	65.3	0.3333393	-0.21
conf_int38	67.51429	0.059954361	56	66.1	67.7	0.3106004	-0.07
conf_int39	68.69123	0.069359045	57	67.8	69.3	0.3024461	-0.06
conf_int40	70.60250	0.150458050	40	69.4	71.7	0.3658329	-0.01
conf_int41	72.85472	-0.177222808	53	71.8	73.9	0.3116525	-0.31
conf_int42	75.71091	0.012589373	55	74.6	76.3	0.3137005	-0.12
conf_int43	77.54902	-0.056629634	51	76.7	78.3	0.3262849	-0.19
conf_int44	80.60811	0.228797075	74	79.3	80.9	0.2668726	0.111
conf_int45	81.63125	-0.396034864	16	81.0	82.4	0.5574244	-0.63
conf_int46	82.96727	0.039342766	55	82.5	84.0	0.3110093	-0.09
conf_int47	84.20000	0.252302313	47	84.2	84.2	0.3368999	0.105
conf_int48	84.98679	-0.332185921	53	84.5	86.5	0.3031489	-0.46
conf_int49	88.59310	-0.382812848	58	86.6	91.1	0.2790762	-0.50
conf_int50	96.64146	-0.059629935	41	91.9	99.9	0.3666793	-0.21

A binned_residuals: 50 × 9

	xbar	ybar	n	x.lo	x.hi	se	CI_l
	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl
conf_int	19.29508	0.252699356	61	18	20	0.2796780	0.130
conf_int1	21.57971	-0.082685721	69	21	22	0.2806257	-0.20
conf_int2	23.00000	0.134071722	47	23	23	0.3249980	-0.01
conf_int3	24.00000	-0.077045204	32	24	24	0.4203514	-0.25
conf_int4	25.00000	-0.050310988	49	25	25	0.3286766	-0.19
conf_int5	26.57778	0.080751232	90	26	27	0.2402476	-0.02
conf_int6	28.00000	0.236950312	40	28	28	0.3484210	0.081
conf_int7	29.00000	0.263383294	53	29	29	0.3013667	0.129
conf_int8	30.00000	-0.260353855	51	30	30	0.3264432	-0.40
conf_int9	31.00000	-0.122194103	49	31	31	0.3290419	-0.26
conf_int10	32.00000	0.155177646	40	32	32	0.3598777	-0.00
conf_int11	33.00000	0.020793220	48	33	33	0.3356858	-0.12
conf_int12	34.00000	0.059707321	48	34	34	0.3334760	-0.08
conf_int13	35.00000	0.141749907	52	35	35	0.3196699	0.002
conf_int14	36.00000	-0.179486502	46	36	36	0.3425105	-0.32
conf_int15	37.00000	0.004014122	53	37	37	0.3177028	-0.13
conf_int16	38.00000	0.263346697	42	38	38	0.3431834	0.111
conf_int17	39.52439	-0.023038670	82	39	40	0.2541881	-0.13
conf_int18	41.00000	-0.007929817	31	41	41	0.4204482	-0.19
conf_int19	42.00000	0.277877666	45	42	42	0.3377048	0.131
conf_int20	43.00000	0.053218208	44	43	43	0.3524159	-0.10
conf_int21	44.35821	0.162999851	67	44	45	0.2776105	0.040
conf_int22	46.00000	-0.178925036	38	46	46	0.3723397	-0.34
conf_int23	47.00000	-0.089516208	53	47	47	0.3183979	-0.22
conf_int24	48.00000	-0.132138628	43	48	48	0.3529043	-0.28
conf_int25	49.56962	0.006727989	79	49	50	0.2592525	-0.10
conf_int26	51.00000	-0.053819937	44	51	51	0.3467750	-0.20
conf_int27	52.00000	-0.171101425	56	52	52	0.3096809	-0.30
conf_int28	53.00000	0.084964907	56	53	53	0.3062914	-0.05
conf_int29	54.00000	-0.087942228	49	54	54	0.3370405	-0.23
conf_int30	55.00000	-0.129463967	45	55	55	0.3536815	-0.28
conf_int31	56.00000	0.009699976	43	56	56	0.3538465	-0.14
conf_int32	57.00000	0.140458736	67	57	57	0.2812257	0.016
conf_int33	58.00000	0.122476280	61	58	58	0.2974261	-0.00
conf_int34	59.00000	0.083529266	55	59	59	0.3093207	-0.05
conf_int35	60.51579	-0.097434209	95	60	61	0.2375947	-0.20
conf_int36	62.00000	0.090866282	53	62	62	0.3195544	-0.04
conf_int37	63.00000	0.036679437	38	63	63	0.3700293	-0.12
conf_int38	64.00000	-0.009923302	30	64	64	0.4293501	-0.19
conf_int39	65.00000	0.389914821	39	65	65	0.3533468	0.234
conf_int40	66.50549	-0.057377378	91	66	67	0.2433667	-0.16
conf_int41	68.00000	-0.198354956	39	68	68	0.3608844	-0.36
conf_int42	69.00000	0.058927115	38	69	69	0.3780426	-0.10
conf_int43	70.45283	-0.016005957	53	70	71	0.3170405	-0.15
conf_int44	72.45238	-0.093241357	42	72	73	0.3652911	-0.25
conf_int45	75.05000	-0.208274842	60	74	76	0.2960546	-0.33
conf_int46	77.55556	0.145327376	45	77	78	0.3356981	-0.00
conf_int47	79.59459	0.127142699	37	79	80	0.3801455	-0.04
conf_int48	82.42373	-0.094142705	59	81	84	0.3001413	-0.22
conf_int49	87.09302	0.134854090	43	85	89	0.3459166	-0.02


Binned Residuals
Points should be within error bounds
Average residual
0.8
0.4
0.0
-0.4
sei10
0
25
50
75
100
Within error bounds
no
yes



Age seems to stick out. It is a better predictor than SES. We may need to add interactions; or add other predictors, like 'polviews', which according to the above model comparisons makes the model significantly better.

[31]:
```
emmeans(fit_3, "polviews") %>% pairs() %>% as.data.frame() %>% filter(p.value <
↪.05)

emmeans(fit_3, "polviews", type="response") %>% pairs() %>% as.data.frame() %>%
↪filter(p.value < .05)
```

A summary_emm: 9 × 6

contrast <chr>	estimate <dbl>	SE <dbl>	df <dbl>	z.ratio <dbl>
Extremely liberal - Moderate	-0.9266262	0.1950664	Inf	-4.750312
Extremely liberal - Slghtly conservative	-0.8487137	0.2127293	Inf	-3.989642
Extremely liberal - Conservative	-0.9935486	0.2108369	Inf	-4.712403
Extremely liberal - Extrmly conservative	-1.3402621	0.2792876	Inf	-4.798860
Liberal - Moderate	-0.7090022	0.1308520	Inf	-5.418353
Liberal - Slghtly conservative	-0.6310897	0.1555805	Inf	-4.056356
Liberal - Conservative	-0.7759246	0.1532082	Inf	-5.064512
Liberal - Extrmly conservative	-1.1226380	0.2392048	Inf	-4.693210
Slightly liberal - Extrmly conservative	-0.7334002	0.2412625	Inf	-3.039843

A summary_emm: 9 × 7

contrast <fct>	odds.ratio <dbl>	SE <dbl>	df <dbl>	null <dbl>
Extremely liberal / Moderate	0.3958871	0.07722426	Inf	1
Extremely liberal / Slghtly conservative	0.4279651	0.09104070	Inf	1
Extremely liberal / Conservative	0.3702605	0.07806458	Inf	1
Extremely liberal / Extrmly conservative	0.2617771	0.07311109	Inf	1
Liberal / Moderate	0.4921350	0.06439684	Inf	1
Liberal / Slghtly conservative	0.5320118	0.08277063	Inf	1
Liberal / Conservative	0.4602780	0.07051835	Inf	1
Liberal / Extrmly conservative	0.3254202	0.07784206	Inf	1
Slightly liberal / Extrmly conservative	0.4802732	0.11587191	Inf	1

Conservatives are 1/0.37 and 1/0.46 times more likely to support mass transit spending compared to extremely liberal and liberal.

Extreme liberals are 0.37, 0.39, and 0.42 times more likely to support spending compared to conservatives, moderates and slight conservatives.

Extrm conservatives are 1/0.32 and 1/0.48 times more likely to support mass spending than liberals and slight liberals.

Liberals are 0.49 and 0.53 times more likely to support spending than moderates and slight conservatives.

10 conclusions

[32]:
```
fit_anova = anova(fit_3, test="Chisq")

fit_anova %>%
  kable()
```

```
|         | Df|  Deviance| Resid. Df| Resid. Dev|  Pr(>Chi)|
|:--------|--:|---------:|---------:|----------:|---------:|
|NULL     | NA|        NA|      2589|   3581.340|        NA|
|age      |  1|  9.268443|      2588|   3572.072| 0.0023314|
|sex      |  1| 12.156624|      2587|   3559.915| 0.0004891|
```

```
|sei10    |  1| 14.119078|       2586|   3545.796| 0.0001716|
|polviews |  6| 63.028441|       2580|   3482.768| 0.0000000|
```

The model best predicts satisfaction of Mass transportation when it includes age, sex, ses and polviews and predictors. Specifically, conservatives are happier in general to mass transportation than liberals; females are happier towards mass transportation than males; and people of lower socioeconomic status are happier towards mass transportation than people of higher socioeconomic status.