Three models for willingness to accept payment

(chr, fct, num)

```
library(tidyverse)
library(chibukuplasticscollection)
```

Data preparation

- Prepare wta variable which has a positive linear relationship with income (monthly_earn1)

```
survey_model <- survey |>
  select(id, age, gender, monthly_earn1) |>
  # positive linear relationship for wta: with higher income, willingness to
  # accept payment increases
  mutate(wta = case_when(
    monthly_earn1 == "Less than MWK25,000" ~ 5,
    monthly_earn1 == "Between MWK25,000 to MWK49,999" ~ 10,
    monthly_earn1 == "Between MWK50,000 to MWK74,999" ~ 20,
    monthly_earn1 == "Between MWK75,000 to MWK99,999" ~ 40,
    monthly_earn1 == "Between MWK100,000 to MWK124,999" ~ 60,
    monthly_earn1 == "Between MWK125,000 to MWK149,999" ~ 90,
    monthly_earn1 == "Between MWK150,000 to MWK174,999" ~ 120,
    monthly_earn1 == "Between MWK175,000 to MWK200,000" ~ 160,
    monthly_earn1 == "More than MWK200,000" ~ 220
  ))

# write csv
survey_model |>
  write_csv(here::here("data/intermediate/survey_small_wta_model.csv"))
```

Data transformation

- nearly 50% of responses cannot be used (I don't know, prefer not to say, NA)
- why so many NAs (20%) ?

```
survey_model |>
  count(monthly_earn1) |>
  mutate(percent = n / sum(n) * 100) |>
  knitr::kable()
```

monthly_earn1	n	percent
Between MWK100,000 to MWK124,999	35	6.317690
Between MWK125,000 to MWK149,999	22	3.971119
Between MWK150,000 to MWK174,999	40	7.220217
Between MWK175,000 to MWK200,000	16	2.888087
Between MWK25,000 to MWK49,999	31	5.595668
Between MWK50,000 to MWK74,999	31	5.595668
Between MWK75,000 to MWK99,999	35	6.317690
I don't know	59	10.649819
Less than MWK25,000	11	1.985560
More than MWK200,000	65	11.732852
Prefer not to say	96	17.328520
NA	113	20.397112

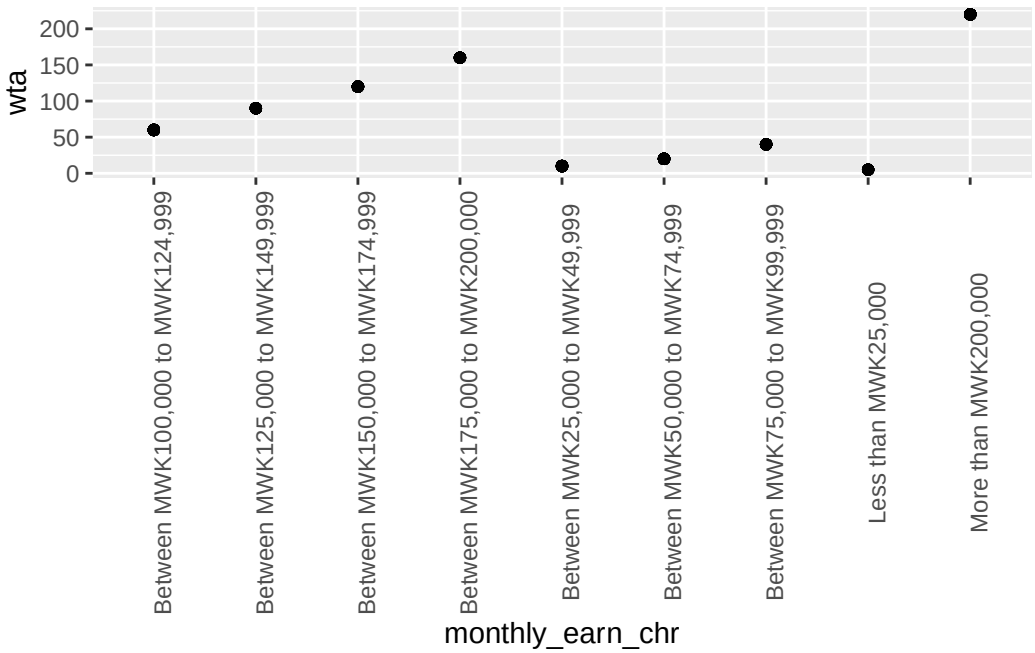
```
# define levels for factor
levels_monthly_earn <- c(
  "Less than MWK25,000",
  "Between MWK25,000 to MWK49,999",
  "Between MWK50,000 to MWK74,999",
  "Between MWK75,000 to MWK99,999",
  "Between MWK100,000 to MWK124,999",
  "Between MWK125,000 to MWK149,999",
  "Between MWK150,000 to MWK174,999",
  "Between MWK175,000 to MWK200,000",
  "More than MWK200,000"
)

survey_model_clean <- survey_model |>
  filter(!monthly_earn1 %in% c("Prefer not to say", "I don't know")) |>
  filter(!is.na(monthly_earn1)) |>
```

```
  rename(monthly_earn_chr = monthly_earn1) |>
  mutate(monthly_earn_fct = fct_relevel(monthly_earn_chr, levels_monthly_earn)) |>
  mutate(monthly_earn_num = case_when(
    monthly_earn_chr == "Less than MWK25,000" ~ 12500,
    monthly_earn_chr == "Between MWK25,000 to MWK49,999" ~ 37500,
    monthly_earn_chr == "Between MWK50,000 to MWK74,999" ~ 62500,
    monthly_earn_chr == "Between MWK75,000 to MWK99,999" ~ 87500,
    monthly_earn_chr == "Between MWK100,000 to MWK124,999" ~ 112500,
    monthly_earn_chr == "Between MWK125,000 to MWK149,999" ~ 137500,
    monthly_earn_chr == "Between MWK150,000 to MWK174,999" ~ 162500,
    monthly_earn_chr == "Between MWK175,000 to MWK200,000" ~ 187500,
    monthly_earn_chr == "More than MWK200,000" ~ 225000
  ))
```

Data visualisation & modeling

Income not ordered (i.e. alphabetically)

```
survey_model_clean |>
  ggplot(aes(x = monthly_earn_chr, y = wta)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90))
```




```r
model_chr_fit <- lm(wta ~ monthly_earn_chr, data = survey_model_clean)
```

Model summary

```r
summary(model_chr_fit)
```

```
Call:
lm(formula = wta ~ monthly_earn_chr, data = survey_model_clean)

Residuals:
       Min         1Q     Median         3Q        Max
-1.626e-12 -3.745e-14  0.000e+00  0.000e+00  2.868e-12

Coefficients:
                                                 Estimate Std. Error
(Intercept)                                     6.000e+01  4.543e-14
monthly_earn_chrBetween MWK125,000 to MWK149,999  3.000e+01  7.313e-14
monthly_earn_chrBetween MWK150,000 to MWK174,999  6.000e+01  6.221e-14
monthly_earn_chrBetween MWK175,000 to MWK200,000  1.000e+02  8.112e-14
monthly_earn_chrBetween MWK25,000 to MWK49,999   -5.000e+01  6.629e-14
monthly_earn_chrBetween MWK50,000 to MWK74,999   -4.000e+01  6.629e-14
```

```
monthly_earn_chrBetween MWK75,000 to MWK99,999   -2.000e+01  6.425e-14
monthly_earn_chrLess than MWK25,000              -5.500e+01  9.291e-14
monthly_earn_chrMore than MWK200,000              1.600e+02  5.635e-14
                                                    t value Pr(>|t|)
(Intercept)                                       1.321e+15   <2e-16 ***
monthly_earn_chrBetween MWK125,000 to MWK149,999  4.102e+14   <2e-16 ***
monthly_earn_chrBetween MWK150,000 to MWK174,999  9.644e+14   <2e-16 ***
monthly_earn_chrBetween MWK175,000 to MWK200,000  1.233e+15   <2e-16 ***
monthly_earn_chrBetween MWK25,000 to MWK49,999   -7.542e+14   <2e-16 ***
monthly_earn_chrBetween MWK50,000 to MWK74,999   -6.034e+14   <2e-16 ***
monthly_earn_chrBetween MWK75,000 to MWK99,999   -3.113e+14   <2e-16 ***
monthly_earn_chrLess than MWK25,000              -5.920e+14   <2e-16 ***
monthly_earn_chrMore than MWK200,000              2.839e+15   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.688e-13 on 277 degrees of freedom
Multiple R-squared:      1, Adjusted R-squared:      1
F-statistic: 3.013e+30 on 8 and 277 DF,  p-value: < 2.2e-16
```

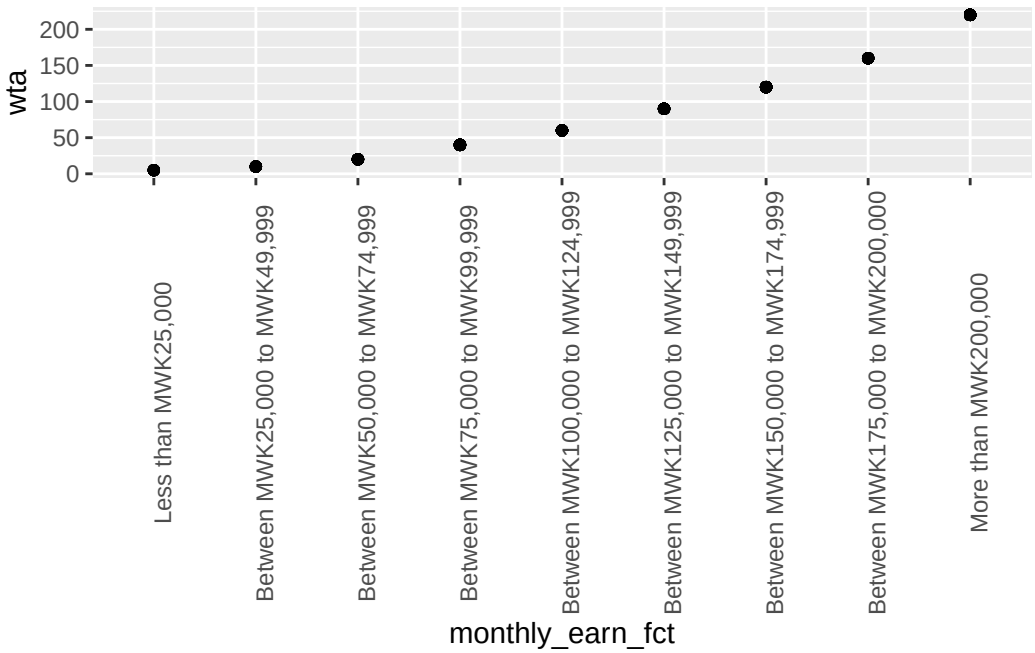
Model coefficients

```
broom::tidy(model_chr_fit)
```

```
# A tibble: 9 x 5
  term                                    estimate std.error statistic p.value
  <chr>                                      <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)                                 60.0  4.54e-14   1.32e15       0
2 monthly_earn_chrBetween MWK125,000 to MW~   30.0  7.31e-14   4.10e14       0
3 monthly_earn_chrBetween MWK150,000 to MW~   60.0  6.22e-14   9.64e14       0
4 monthly_earn_chrBetween MWK175,000 to MW~  100.   8.11e-14   1.23e15       0
5 monthly_earn_chrBetween MWK25,000 to MWK~  -50.0  6.63e-14  -7.54e14       0
6 monthly_earn_chrBetween MWK50,000 to MWK~  -40.0  6.63e-14  -6.03e14       0
7 monthly_earn_chrBetween MWK75,000 to MWK~  -20.0  6.43e-14  -3.11e14       0
8 monthly_earn_chrLess than MWK25,000        -55.0  9.29e-14  -5.92e14       0
9 monthly_earn_chrMore than MWK200,000       160.   5.64e-14   2.84e15       0
```

Income ordered

```
survey_model_clean |>
  ggplot(aes(x = monthly_earn_fct, y = wta)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90))
```




```
model_fct_fit <- lm(wta ~ monthly_earn_fct, data = survey_model_clean)
```

Model summary

```
summary(model_fct_fit)
```

```
Call:
lm(formula = wta ~ monthly_earn_fct, data = survey_model_clean)

Residuals:
       Min          1Q      Median          3Q         Max
-2.480e-12  0.000e+00  0.000e+00  3.517e-14  1.921e-12
```

```
Coefficients:
                                                 Estimate Std. Error   t value
(Intercept)                                     5.000e+00  7.250e-14 6.896e+13
monthly_earn_fctBetween MWK25,000 to MWK49,999  5.000e+00  8.439e-14 5.925e+13
monthly_earn_fctBetween MWK50,000 to MWK74,999  1.500e+01  8.439e-14 1.777e+14
monthly_earn_fctBetween MWK75,000 to MWK99,999  3.500e+01  8.312e-14 4.211e+14
monthly_earn_fctBetween MWK100,000 to MWK124,999 5.500e+01  8.312e-14 6.617e+14
monthly_earn_fctBetween MWK125,000 to MWK149,999 8.500e+01  8.880e-14 9.572e+14
monthly_earn_fctBetween MWK150,000 to MWK174,999 1.150e+02  8.187e-14 1.405e+15
monthly_earn_fctBetween MWK175,000 to MWK200,000 1.550e+02  9.419e-14 1.646e+15
monthly_earn_fctMore than MWK200,000            2.150e+02  7.840e-14 2.742e+15
                                                Pr(>|t|)
(Intercept)                                       <2e-16 ***
monthly_earn_fctBetween MWK25,000 to MWK49,999    <2e-16 ***
monthly_earn_fctBetween MWK50,000 to MWK74,999    <2e-16 ***
monthly_earn_fctBetween MWK75,000 to MWK99,999    <2e-16 ***
monthly_earn_fctBetween MWK100,000 to MWK124,999  <2e-16 ***
monthly_earn_fctBetween MWK125,000 to MWK149,999  <2e-16 ***
monthly_earn_fctBetween MWK150,000 to MWK174,999  <2e-16 ***
monthly_earn_fctBetween MWK175,000 to MWK200,000  <2e-16 ***
monthly_earn_fctMore than MWK200,000              <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.405e-13 on 277 degrees of freedom
Multiple R-squared:      1, Adjusted R-squared:      1
F-statistic: 3.765e+30 on 8 and 277 DF,  p-value: < 2.2e-16
```

Model coefficients

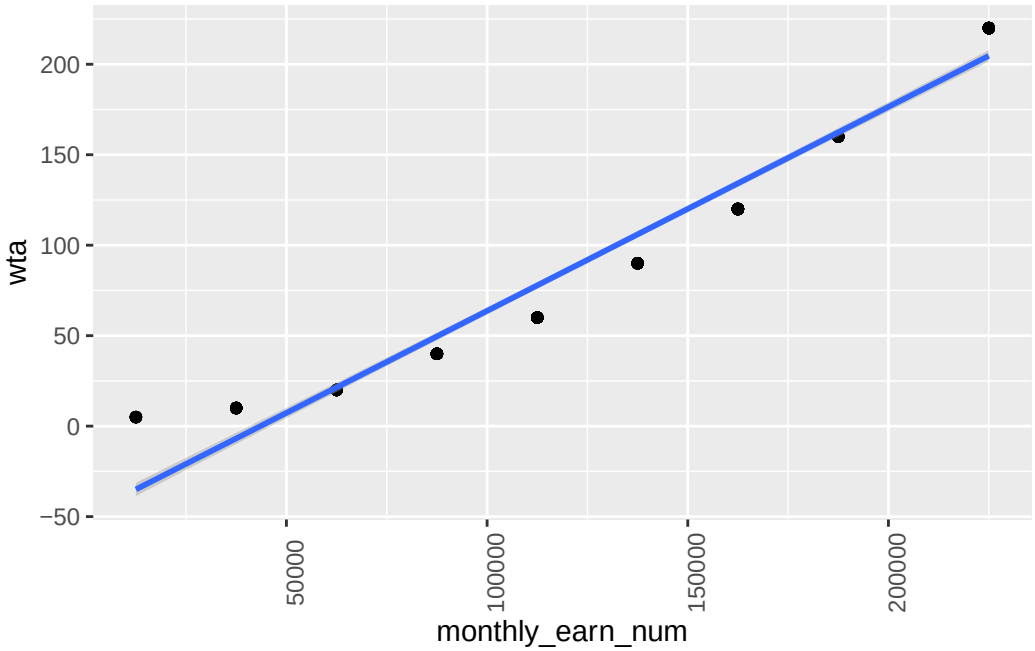
```
broom::tidy(model_fct_fit)
```

```
# A tibble: 9 x 5
  term                                    estimate std.error statistic p.value
  <chr>                                      <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)                                 5.00  7.25e-14   6.90e13       0
2 monthly_earn_fctBetween MWK25,000 to MWK~    5.00  8.44e-14   5.92e13       0
3 monthly_earn_fctBetween MWK50,000 to MWK~   15.0   8.44e-14   1.78e14       0
4 monthly_earn_fctBetween MWK75,000 to MWK~   35.0   8.31e-14   4.21e14       0
5 monthly_earn_fctBetween MWK100,000 to MW~   55.0   8.31e-14   6.62e14       0
6 monthly_earn_fctBetween MWK125,000 to MW~   85.0   8.88e-14   9.57e14       0
7 monthly_earn_fctBetween MWK150,000 to MW~  115     8.19e-14   1.40e15       0
```

```
8 monthly_earn_fctBetween MWK175,000 to MW~   155.    9.42e-14   1.65e15       0
9 monthly_earn_fctMore than MWK200,000        215     7.84e-14   2.74e15       0
```

Income as numeric

```
survey_model_clean |>
  ggplot(aes(x = monthly_earn_num, y = wta)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme(axis.text.x = element_text(angle = 90))
```




```
model_num_fit <- lm(wta ~ monthly_earn_num, data = survey_model_clean)
```

Model summary

```
summary(model_num_fit)
```

```
Call:
lm(formula = wta ~ monthly_earn_num, data = survey_model_clean)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-17.789 -14.167  -2.355  15.361  39.966

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -4.906e+01  2.012e+00  -24.38   <2e-16 ***
monthly_earn_num  1.128e-03  1.366e-05   82.52   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.67 on 284 degrees of freedom
Multiple R-squared:   0.96, Adjusted R-squared:  0.9598
F-statistic:  6810 on 1 and 284 DF,  p-value: < 2.2e-16
```

Model coefficients

```
broom::tidy(model_num_fit)
```

```
# A tibble: 2 x 5
  term              estimate std.error statistic   p.value
  <chr>                <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)      -49.1     2.01          -24.4 1.32e- 71
2 monthly_earn_num   0.00113 0.0000137      82.5 1.69e-200
```