# Mixed Effect Modeling

Load Data

```r
# Regular data
homelessness_data = read.csv('../data/processed/pivoted_and_PIT.csv')

#Standardized Data
standardize = function(col){
  col = (col - mean(col)) / sd(col)
  return(col)
}

homelessness_data_standard = homelessness_data |>
  mutate(across(where(is.numeric) & !all_of("Unsheltered.Per.100.000"), standardize))
```

Why Mixed Effects Model

```r
# Examine variation in Unsheltered.Per.100.000 by City
ggplot(homelessness_data, aes(x = City, y = Unsheltered.Per.100.000)) +
  geom_boxplot() +
  labs(title = "Distribution of Unsheltered Rate per 100,000 by City") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Distribution of Unsheltered Rate per 100,000 by City
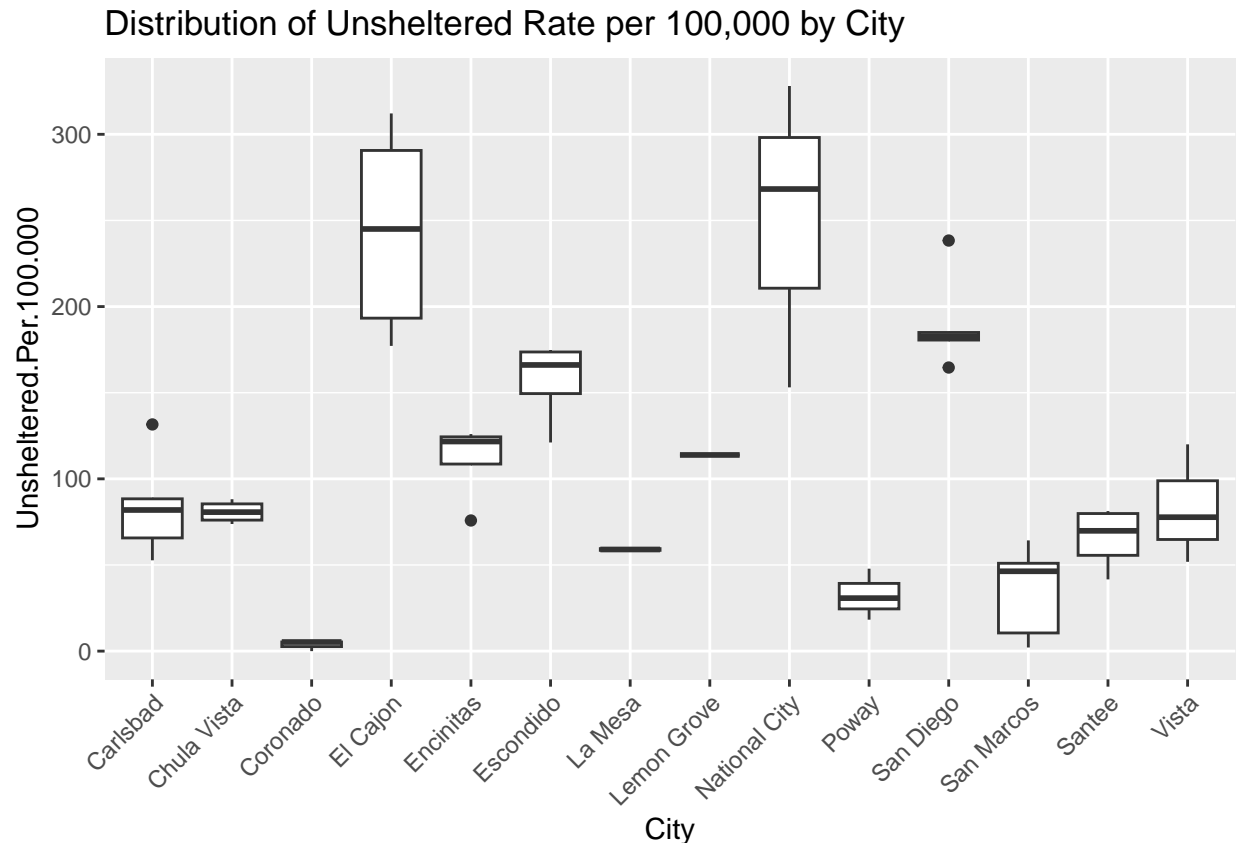
This graph shows variability based on city, and if we want to focus on effects of programs on unsheltered rather than city influences on those programs, it would make sense to treat city as a random effect

Implement Models

```r
ignored_features = c("PEH.Per.100.000", "Total.PEH", "Unsheltered.PEH", "Population")

model <- lmer(Unsheltered.Per.100.000 ~
            bridge.to.housing.network + emergency.shelter + family.reunification.program +
            flexible.funds + food.and.nutrition + homeless.services + homelessness.prevention +
            homeshare.program + housing.assistance + housing.navigation.services +
            housing.stability.services + motel.voucher + neighborhood.revitalization.services +
            opening.doors.program + outreach + project.h.o.p.e. + rapid.re.housing +
            rental.assistance + restrooms + safe.parking + service.center + staff.and.operations +
            take.back.the.streets + transitional.housing + work.for.hope +
            (1  | City), data = homelessness_data |> select(-all_of(ignored_features)))
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```
#Standarized model
modelS <- lmer(Unsheltered.Per.100.000 ~
               bridge.to.housing.network + emergency.shelter + family.reunification.program +
               flexible.funds + food.and.nutrition + homeless.services + homelessness.prevention +
               homeshare.program + housing.assistance + housing.navigation.services +
               housing.stability.services + motel.voucher + neighborhood.revitalization.services +
               opening.doors.program + outreach + project.h.o.p.e. + rapid.re.housing +
               rental.assistance + restrooms + safe.parking + service.center + staff.and.operations +
               take.back.the.streets + transitional.housing + work.for.hope +
               (1 | City), data = homelessness_data_standard |> select(-all_of(ignored_features)))
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

Addressing Warnings

```
vif(model)
```

```
##                                          GVIF Df GVIF^(1/(2*Df))
## bridge.to.housing.network            2.119637e+01  1        4.603952
## emergency.shelter                    3.019310e+01  1        5.494825
## family.reunification.program         2.494188e+01  1        4.994185
## flexible.funds                       1.984565e+02  1       14.087458
## food.and.nutrition                   1.265858e+00  1        1.125104
## homeless.services                    5.727109e+01  1        7.567767
## homelessness.prevention              1.368564e+00  1        1.169856
## homeshare.program                    1.279381e+00  1        1.131097
## housing.assistance                   6.251358e+00  1        2.500272
## housing.navigation.services          3.900159e+01  1        6.245125
## housing.stability.services           1.993901e+02  1       14.120556
## motel.voucher                        2.016420e+01  1        4.490456
## neighborhood.revitalization.services 1.312766e+00  1        1.145760
## opening.doors.program                3.907793e+00  1        1.976814
## outreach                             8.250283e+01  1        9.083107
## project.h.o.p.e.                     3.118510e+00  1        1.765930
## rapid.re.housing                     4.773277e+01  1        6.908891
## rental.assistance                    1.401692e+00  1        1.183931
## restrooms                            5.898827e+01  1        7.680382
## safe.parking                         6.839702e+01  1        8.270249
## service.center                       3.003647e+01  1        5.480554
## staff.and.operations                 1.907766e+00  1        1.381219
## take.back.the.streets                2.182301e+00  1        1.477261
## transitional.housing                 2.416204e+01  1        4.915490
## work.for.hope                        3.644808e+13  0             Inf
```

The VIF for work.for.hope is extremely large (4.08e+13) with an infinite value for the scaled GVIF, indicating severe multicollinearity, likely because this variable is almost perfectly collinear with one or more other predictors in the model. This is the source of the "rank deficient" warning, and R removes work.for.hope from the model.

Visualizing City-Level Random Effects to Justify Mixed-Effects Model Usage

```
# Random effects for City
ranef_city <- ranef(model)$City
ggplot(ranef_city, aes(x = reorder(row.names(ranef_city), `(Intercept)`), y = `(Intercept)`)) +
  geom_point() +
  coord_flip() +
  labs(title = "City-Level Random Effects", x = "City", y = "Random Intercept Estimate")
```

City–Level Random Effects
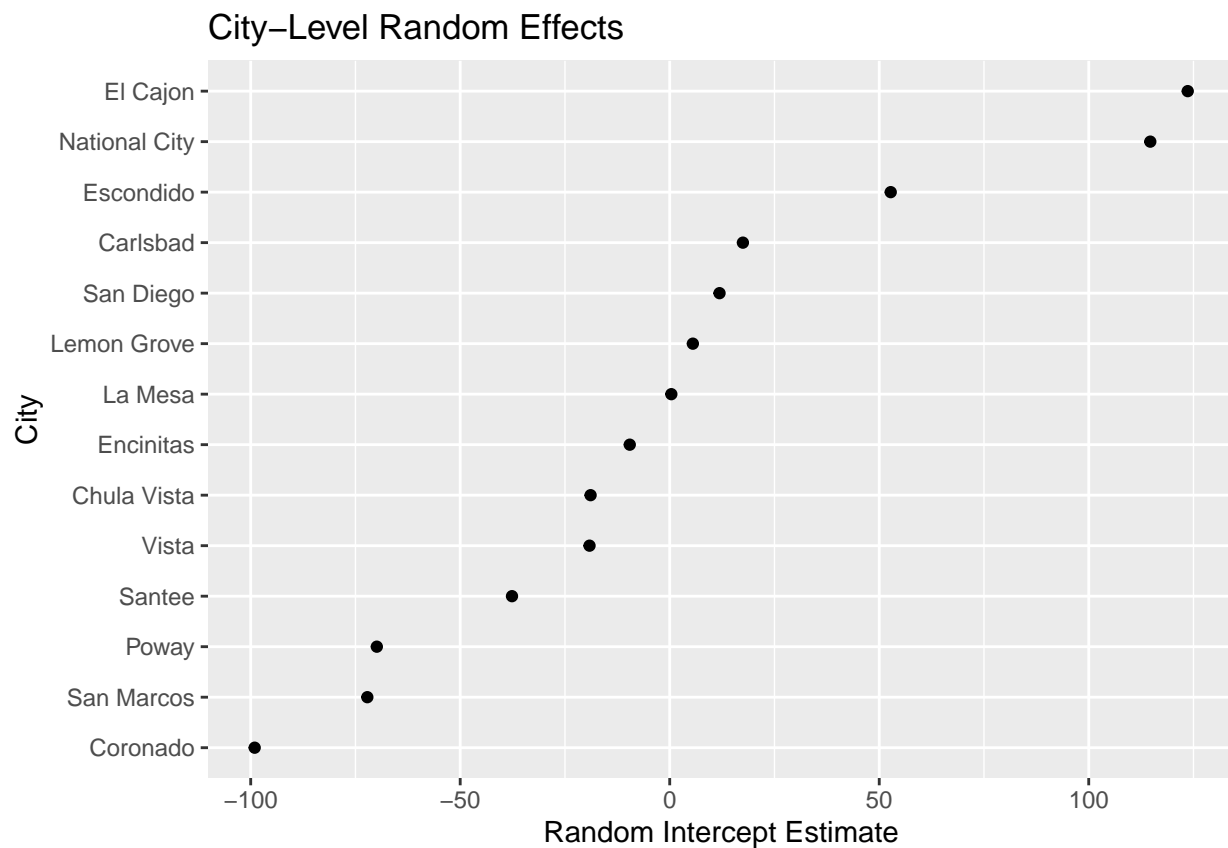
Model Summaries

```
# For PDF
summarize_model <- function(model) {
  # Extract the full model summary
  model_summary <- summary(model)

  # Convert the fixed effects table to a data frame
  fixed_effects <- as.data.frame(model_summary$coefficients)

  # Sort the fixed effects by Estimate
```

```r
  sorted_fixed_effects <- fixed_effects[order(fixed_effects$Estimate), ]

  # Add the C-like format for estimates, but remove the predictor names in the formatted column
  sorted_fixed_effects$Estimates <- sprintf("%.4f", sorted_fixed_effects$Estimate)

  # Remove the original 'Estimate' column
  sorted_fixed_effects <- sorted_fixed_effects[, !names(sorted_fixed_effects) %in% c("Estimate")]

  # Rearrange the columns to move 'Estimates' to the front
  sorted_fixed_effects <- sorted_fixed_effects[, c("Estimates", setdiff(names(sorted_fixed_effects), "E

  # Print the modified summary with formatted estimates
  sorted_fixed_effects %>%
    kable("latex", booktabs = TRUE) %>%
    kable_styling(latex_options = "scale_down")
}

# For Console
# summarize_model <- function(model) {
#    # Extract the full model summary
#    model_summary <- summary(model)
#
#    # Convert the fixed effects table to a data frame
#    fixed_effects <- as.data.frame(model_summary$coefficients)
#
#    # Sort the fixed effects by Estimate
#    sorted_fixed_effects <- fixed_effects[order(fixed_effects$Estimate), ]
#
#    # Add the C-like format for estimates, but remove the predictor names in the formatted column
#    sorted_fixed_effects$Estimates <- sprintf("%.4f", sorted_fixed_effects$Estimate)
#
#    # Remove the original 'Estimate' column
#    sorted_fixed_effects <- sorted_fixed_effects[, !names(sorted_fixed_effects) %in% c("Estimate")]
#
#    # Rearrange the columns to move 'Estimates' to the front
#    sorted_fixed_effects <- sorted_fixed_effects[, c("Estimates", setdiff(names(sorted_fixed_effects),
#
#    # Print the modified summary with formatted estimates
#    print(sorted_fixed_effects)
# }

summarize_model(model)


summarize_model(modelS)
```

	Estimates	Std. Error	df	t value	Pr(>|t|)
bridge.to.housing.network	-0.0041	0.0050058	21.39161	-0.8184315	0.4221347
project.h.o.p.e.	-0.0036	0.0069981	22.92499	-0.5093936	0.6153450
flexible.funds	-0.0019	0.0021488	21.57902	-0.8732437	0.3921444
homeshare.program	-0.0013	0.0016799	23.74414	-0.7459848	0.4629987
neighborhood.revitalization.services	-0.0005	0.0004868	27.31544	-1.0885173	0.2858779
service.center	-0.0002	0.0002378	28.61813	-0.9076437	0.3716465
family.reunification.program	-0.0002	0.0002574	30.27835	-0.8201204	0.4185572
take.back.the.streets	-0.0001	0.0003659	21.93376	-0.3378178	0.7387124
homeless.services	-0.0001	0.0008656	23.65704	-0.1426506	0.8877742
rapid.re.housing	-0.0001	0.0001161	29.49783	-1.0324982	0.3102303
outreach	-0.0001	0.0001333	25.03905	-0.6570337	0.5171535
housing.navigation.services	-0.0001	0.0001576	15.10570	-0.5389667	0.5977642
rental.assistance	-0.0000	0.0000452	22.19783	-0.1551795	0.8780822
emergency.shelter	0.0000	0.0000141	30.90048	1.4946802	0.1451410
housing.assistance	0.0000	0.0000365	23.30938	0.6188417	0.5420219
food.and.nutrition	0.0000	0.0003072	22.96635	0.1315442	0.8964904
transitional.housing	0.0001	0.0000590	29.81341	0.8907320	0.3802051
homelessness.prevention	0.0001	0.0003003	25.23164	0.1799794	0.8586064
staff.and.operations	0.0001	0.0001419	24.67271	1.0374071	0.3096070
safe.parking	0.0002	0.0003522	30.53654	0.4395803	0.6633364
opening.doors.program	0.0005	0.0007804	24.59442	0.6363629	0.5304183
motel.voucher	0.0006	0.0012338	24.46260	0.5239744	0.6050118
housing.stability.services	0.0010	0.0012060	21.93836	0.8076378	0.4279697
restrooms	0.0023	0.0030419	23.14728	0.7604965	0.4546353
(Intercept)	108.2278	24.6232458	13.00387	4.3953524	0.0007233

|	Estimates	Std. Error	df	t value	Pr(>|t|)
flexible.funds	-94.7270	108.477125	21.57902	-0.8732437	0.3921444
rapid.re.housing	-48.2851	46.765330	29.49783	-1.0324982	0.3102303
family.reunification.program	-42.3905	51.688103	30.27835	-0.8201204	0.4185572
outreach	-38.3456	58.361741	25.03905	-0.6570337	0.5171535
service.center	-35.7506	39.388335	28.61813	-0.9076437	0.3716465
bridge.to.housing.network	-29.5786	36.140634	21.39161	-0.8184315	0.4221347
housing.navigation.services	-21.8628	40.564354	15.10570	-0.5389667	0.5977642
neighborhood.revitalization.services	-14.0285	12.887736	27.31544	-1.0885173	0.2858779
homeless.services	-7.3345	51.416086	23.65704	-0.1426506	0.8877742
project.h.o.p.e.	-5.9545	11.689483	22.92499	-0.5093936	0.6153450
homeshare.program	-5.8613	7.857138	23.74414	-0.7459848	0.4629987
take.back.the.streets	-3.3034	9.778653	21.93376	-0.3378178	0.7387124
rental.assistance	-1.2498	8.054077	22.19783	-0.1551795	0.8780822
food.and.nutrition	1.0116	7.689997	22.96635	0.1315442	0.8964904
homelessness.prevention	1.3912	7.729939	25.23164	0.1799794	0.8586064
staff.and.operations	9.8517	9.496502	24.67271	1.0374071	0.3096070
housing.assistance	9.9260	16.039611	23.30938	0.6188417	0.5420219
opening.doors.program	13.0016	20.431050	24.59442	0.6363629	0.5304183
motel.voucher	17.0784	32.593932	24.46260	0.5239744	0.6050118
safe.parking	29.6583	67.469696	30.53654	0.4395803	0.6633364
restrooms	39.2690	51.636027	23.14728	0.7604965	0.4546353
transitional.housing	45.1590	50.698709	29.81341	0.8907320	0.3802051
emergency.shelter	74.7368	50.001877	30.90048	1.4946802	0.1451410
housing.stability.services	87.9868	108.943342	21.93836	0.8076378	0.4279697
(Intercept)	110.4264	21.613701	10.48731	5.1090948	0.0003942

Plot Estimates

```r
# Extract fixed effects
fixed_effects <- as.data.frame(summary(model)$coefficients)

# Remove the intercept row
fixed_effects_no_intercept <- fixed_effects[rownames(fixed_effects) != "(Intercept)", ]

# Reorder the predictors based on the Estimate
fixed_effects_no_intercept$Predictor <- factor(
  rownames(fixed_effects_no_intercept),
  levels = rownames(fixed_effects_no_intercept)[order(fixed_effects_no_intercept$Estimate)]
)

# Plot the fixed effects excluding the intercept, sorted by Estimate
ggplot(fixed_effects_no_intercept, aes(x = Predictor, y = Estimate)) +
  geom_bar(stat = "identity") +
  theme(
    axis.text.x = element_text(angle = 60, hjust = 1, size = 10), # Rotate and adjust the labels
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12)
  ) +
  labs(
    title = "Fixed Effects Coefficients",
    x = "Predictors",
    y = "Estimate"
  )
```

Fixed Effects Coefficients
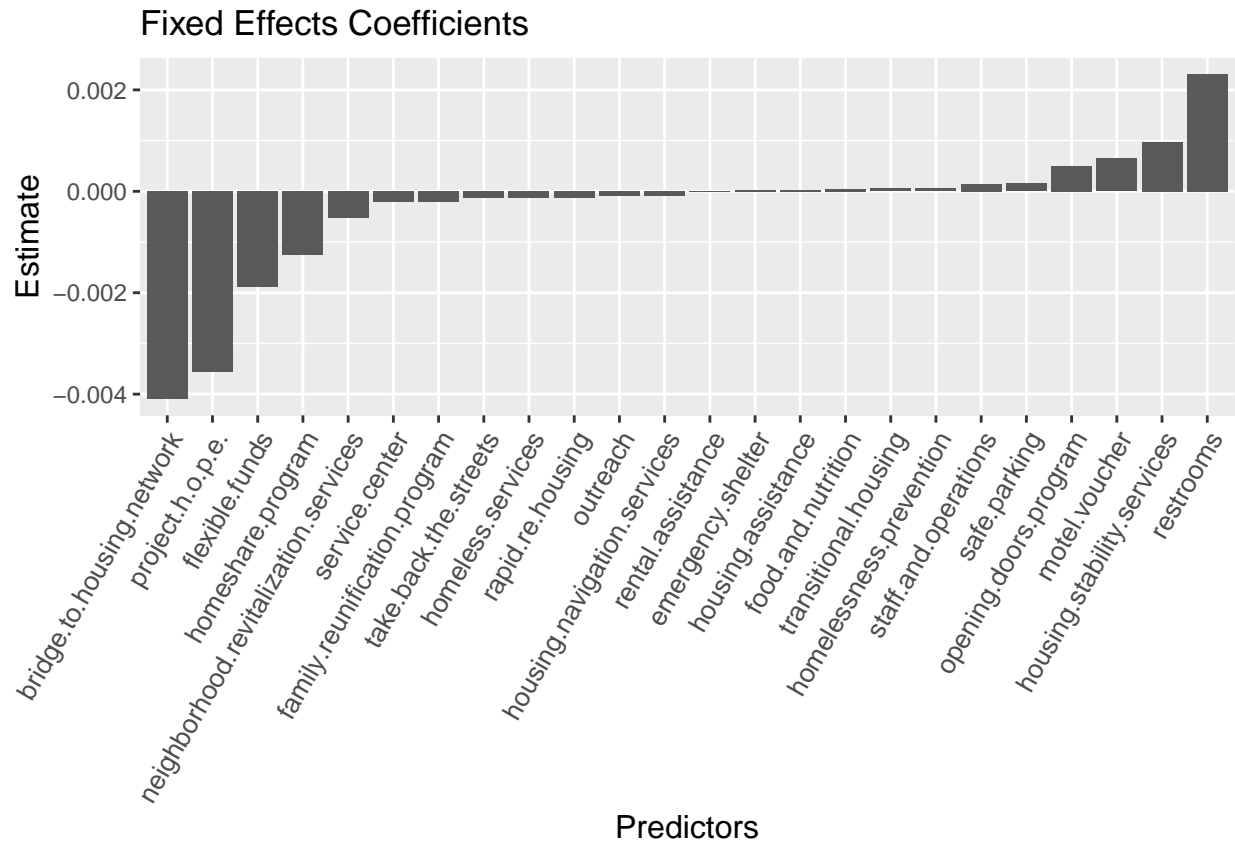

```r
# Extract fixed effects
fixed_effectsS <- as.data.frame(summary(modelS)$coefficients)

# Remove the intercept row
fixed_effects_no_interceptS <- fixed_effectsS[rownames(fixed_effectsS) != "(Intercept)", ]

# Reorder the predictors based on the Estimate
fixed_effects_no_interceptS$Predictor <- factor(
  rownames(fixed_effects_no_interceptS),
  levels = rownames(fixed_effects_no_interceptS)[order(fixed_effects_no_interceptS$Estimate)]
)

# Plot the fixed effects excluding the intercept, sorted by Estimate
ggplot(fixed_effects_no_interceptS, aes(x = Predictor, y = Estimate)) +
  geom_bar(stat = "identity") +
  theme(
    axis.text.x = element_text(angle = 60, hjust = 1, size = 10), # Rotate and adjust the labels
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12)
  ) +
  labs(
    title = "Fixed Effects Coefficients For Standarzied Model",
    x = "Predictors",
    y = "Estimate"
  )
```

Fixed Effects Coefficients For Standarzied Model
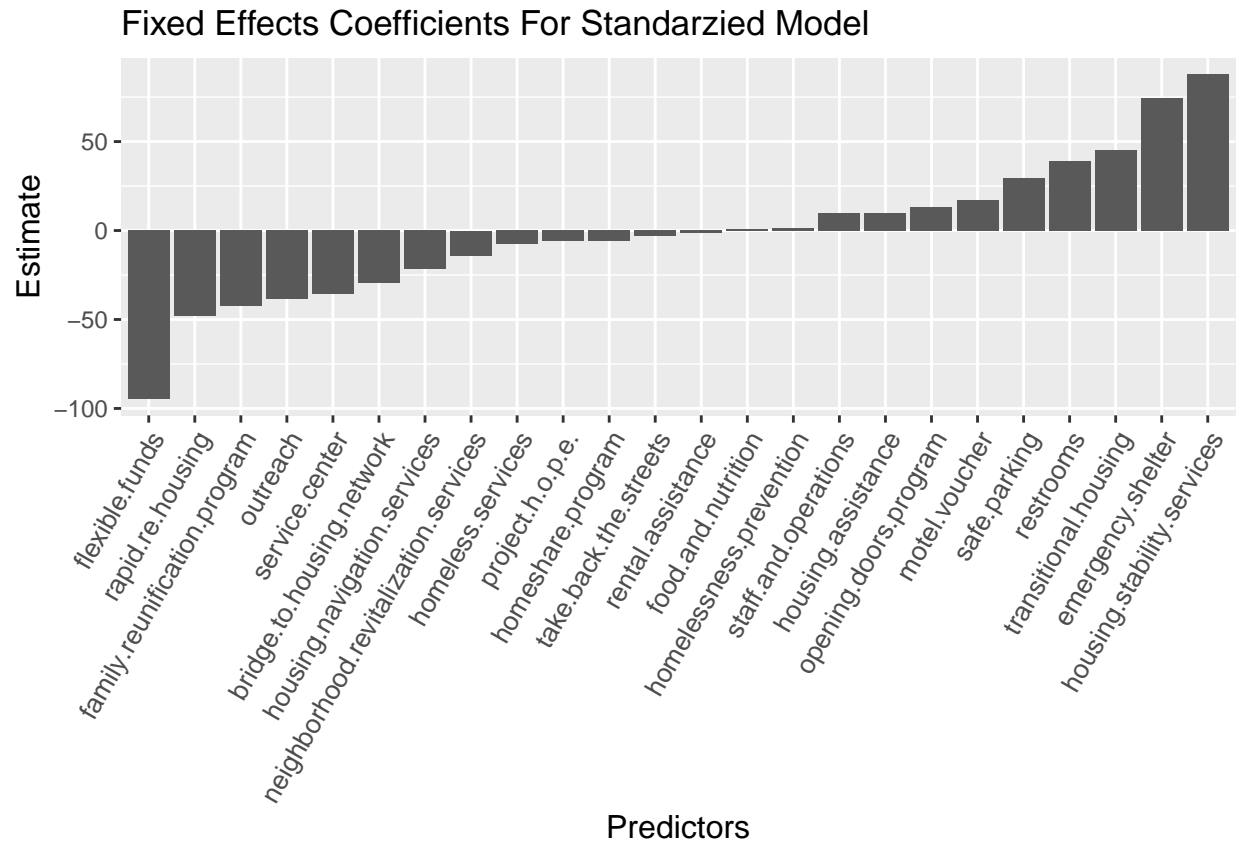
Flexible funds in both graphs as one of lower 3, and housing stability services in both graphs as upper 3. Lower indicates decrease in unsheltered as funds allocated for program, and upper indicates increase.
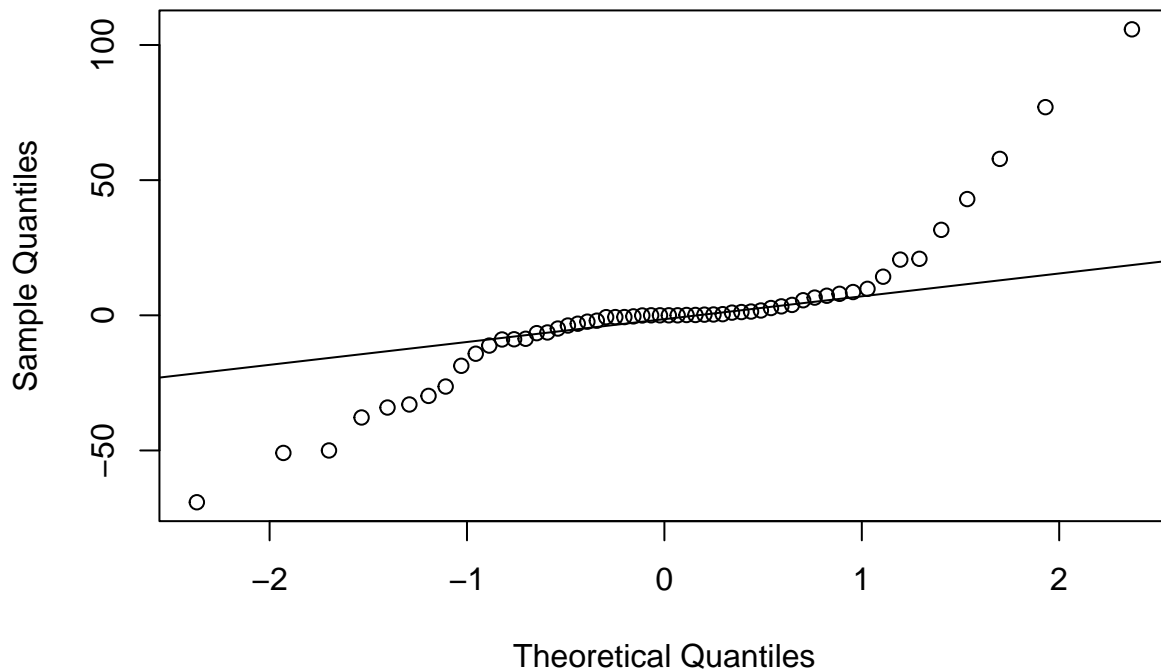
Assessing Residuals and Model Fit

```r
# Plot residuals
plot(resid(model))
```
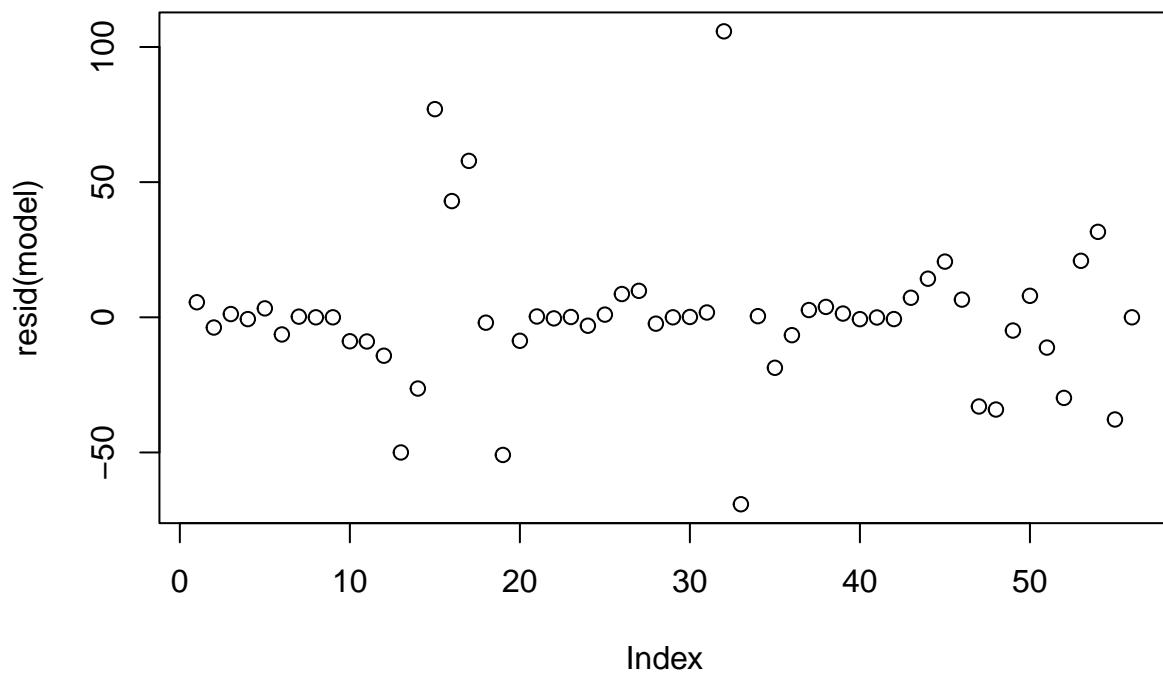
```
qqnorm(resid(model))
qqline(resid(model))
```

Normal Q–Q Plot

Variance in residuals in constant for most part, and no overall patterns which is a good sign. There are some outliers which can be examined

Results if outliers are removed

```r
# Extract residuals from the model
residuals <- resid(model)

# Set a threshold for identifying outliers
# For example, residuals greater than 2 standard deviations from the mean
threshold <- 2 * sd(residuals)

# Identify the indices of the outliers
outlier_indices <- which(abs(residuals) > threshold)

# Display the indices and residual values of the outliers
outliers <- data.frame(Index = outlier_indices, Residual = residuals[outlier_indices])
print(outliers)
```

```
##    Index  Residual
## 15    15  77.00378
## 17    17  57.86662
## 32    32 105.78546
## 33    33 -69.15055
```

	Estimates	Std. Error	df	t value	Pr(>|t|)
bridge.to.housing.network	-0.0041	0.0032795	16.37769	-1.2521319	0.2281088
project.h.o.p.e.	-0.0040	0.0046361	18.31420	-0.8684920	0.3963638
homeshare.program	-0.0020	0.0012874	24.76485	-1.5869303	0.1252158
flexible.funds	-0.0019	0.0014165	16.82570	-1.3721891	0.1880227
neighborhood.revitalization.services	-0.0007	0.0003552	24.30181	-1.9914134	0.0577899
service.center	-0.0003	0.0001642	24.83715	-1.6688620	0.1077038
family.reunification.program	-0.0002	0.0001873	25.04342	-1.1082740	0.2782742
take.back.the.streets	-0.0001	0.0002430	17.46802	-0.5006767	0.6228484
rapid.re.housing	-0.0001	0.0000823	26.51813	-1.3802994	0.1790203
homeless.services	-0.0001	0.0005910	19.52383	-0.1849672	0.8551627
housing.navigation.services	-0.0001	0.0001261	12.07971	-0.5688731	0.5798655
outreach	-0.0000	0.0000920	19.77669	-0.4843383	0.6334663
rental.assistance	-0.0000	0.0000315	20.40753	-0.1264918	0.9005806
safe.parking	0.0000	0.0002536	26.88074	0.0525117	0.9585093
emergency.shelter	0.0000	0.0000101	26.88842	2.0604259	0.0491546
housing.assistance	0.0000	0.0000249	18.80925	1.6689511	0.1116873
transitional.housing	0.0001	0.0000419	26.14695	1.7741930	0.0876782
food.and.nutrition	0.0001	0.0002025	17.18965	0.3930064	0.6991459
homelessness.prevention	0.0002	0.0002050	19.65064	0.7372626	0.4696750
staff.and.operations	0.0003	0.0001519	25.55805	1.9618820	0.0607534
opening.doors.program	0.0007	0.0005440	21.41411	1.2784357	0.2147790
motel.voucher	0.0007	0.0008386	20.95343	0.8603854	0.3993147
housing.stability.services	0.0010	0.0007981	17.37370	1.2827368	0.2164342
restrooms	0.0017	0.0020135	17.65967	0.8633347	0.3995273
(Intercept)	95.0640	21.0178262	11.58061	4.5230182	0.0007629

```r
# View the corresponding rows in the original data
outlier_data <- homelessness_data[outlier_indices, ]

# Remove the identified outlier indices from the dataset
homelessness_data_no_outliers <- homelessness_data[-c(15, 17, 32, 33), ]

# Re-run the model without outliers
model_no_outliers <- lmer(Unsheltered.Per.100.000 ~
            bridge.to.housing.network + emergency.shelter + family.reunification.program +
            flexible.funds + food.and.nutrition + homeless.services + homelessness.prevention +
            homeshare.program + housing.assistance + housing.navigation.services +
            housing.stability.services + motel.voucher + neighborhood.revitalization.services +
            opening.doors.program + outreach + project.h.o.p.e. + rapid.re.housing +
            rental.assistance + restrooms + safe.parking + service.center + staff.and.operations +
            take.back.the.streets + transitional.housing + work.for.hope +
            (1 | City),
            data = homelessness_data_no_outliers |> select(-all_of(ignored_features)))
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```r
summarize_model(model_no_outliers)
```

	Estimates	Std. Error	df	t value	Pr(>|t|)
flexible.funds	-101.8162	74.199805	16.825696	-1.3721891	0.1880227
rapid.re.housing	-47.4330	34.364296	26.518128	-1.3802994	0.1790203
service.center	-47.0326	28.182460	24.837153	-1.6688620	0.1077038
family.reunification.program	-43.1701	38.952566	25.043419	-1.1082740	0.2782742
bridge.to.housing.network	-30.7442	24.553502	16.377690	-1.2521319	0.2281088
outreach	-20.2259	41.759773	19.776689	-0.4843383	0.6334663
neighborhood.revitalization.services	-19.3919	9.737759	24.301811	-1.9914134	0.0577899
housing.navigation.services	-19.1628	33.685575	12.079714	-0.5688731	0.5798655
homeshare.program	-9.9161	6.248606	24.764852	-1.5869303	0.1252158
project.h.o.p.e.	-6.9795	8.036401	18.314204	-0.8684920	0.3963638
homeless.services	-6.7294	36.381526	19.523826	-0.1849672	0.8551627
take.back.the.streets	-3.3750	6.740922	17.468016	-0.5006767	0.6228484
rental.assistance	-0.7355	5.814736	20.407533	-0.1264918	0.9005806
food.and.nutrition	2.0652	5.254896	17.189654	0.3930064	0.6991459
safe.parking	2.6423	50.319106	26.880741	0.0525117	0.9585093
homelessness.prevention	4.0358	5.474011	19.650644	0.7372626	0.4696750
opening.doors.program	18.8671	14.757931	21.414111	1.2784357	0.2147790
housing.assistance	18.9042	11.326979	18.809251	1.6689511	0.1116873
motel.voucher	19.7418	22.945296	20.953429	0.8603854	0.3993147
staff.and.operations	20.5038	10.451089	25.558048	1.9618820	0.0607534
restrooms	30.6133	35.459389	17.659670	0.8633347	0.3995273
transitional.housing	66.1363	37.276858	26.146954	1.7741930	0.0876782
emergency.shelter	76.0686	36.918866	26.888424	2.0604259	0.0491546
housing.stability.services	95.9683	74.815254	17.373696	1.2827368	0.2164342
(Intercept)	105.6778	18.042531	9.264697	5.8571506	0.0002159

```r
#Standardized Model with no outliers
homelessness_data_no_outliers_standard = homelessness_data_no_outliers |>
  mutate(across(where(is.numeric) & !all_of("Unsheltered.Per.100.000"), standardize))

model_no_outliersS <- lmer(Unsheltered.Per.100.000 ~
             bridge.to.housing.network + emergency.shelter + family.reunification.program +
             flexible.funds + food.and.nutrition + homeless.services + homelessness.prevention +
             homeshare.program + housing.assistance + housing.navigation.services +
             housing.stability.services + motel.voucher + neighborhood.revitalization.services +
             opening.doors.program + outreach + project.h.o.p.e. + rapid.re.housing +
             rental.assistance + restrooms + safe.parking + service.center + staff.and.operations +
             take.back.the.streets + transitional.housing + work.for.hope +
             (1 | City),
             data = homelessness_data_no_outliers_standard |> select(-all_of(ignored_features)))
```

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

```r
summarize_model(model_no_outliersS)
```

```r
#Graph estmates for models without ouliers
# Extract fixed effects
fixed_effects <- as.data.frame(summary(model)$coefficients)
```
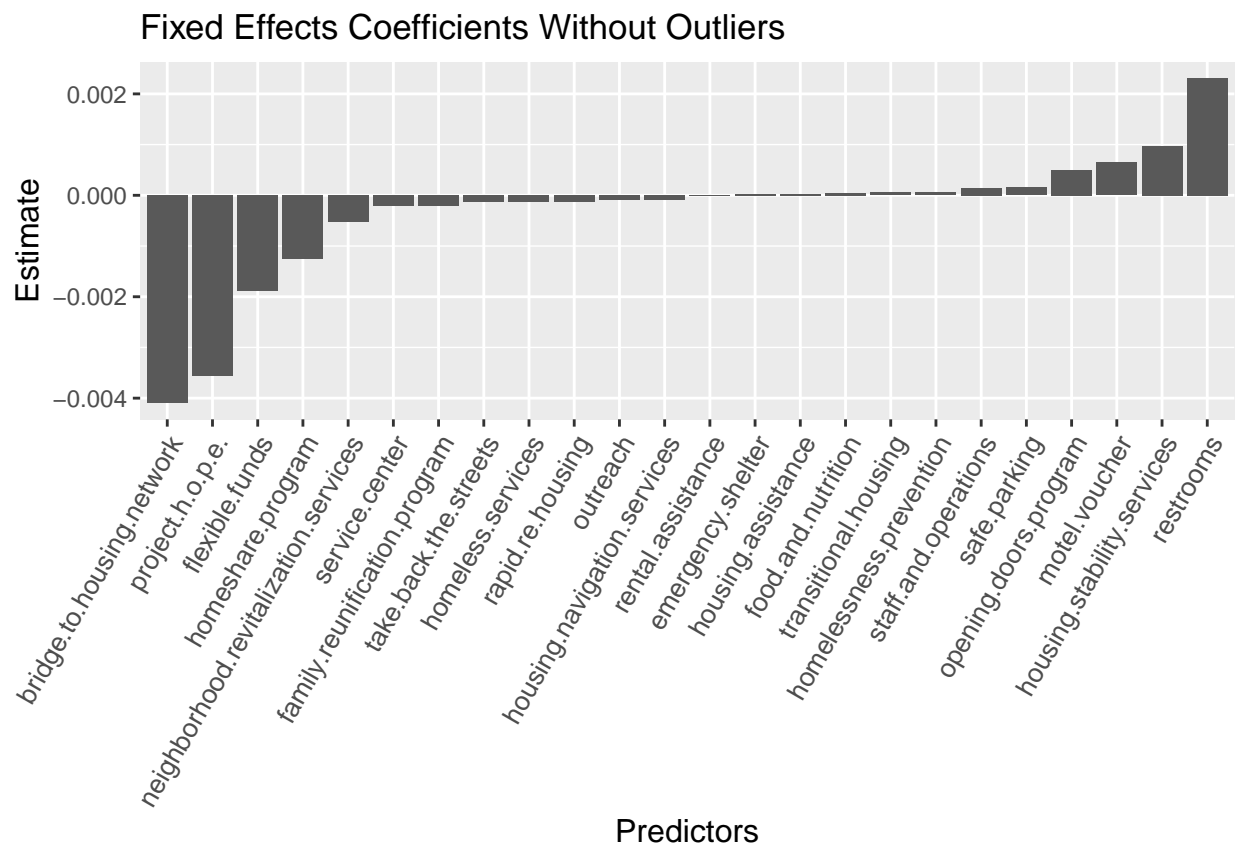
```r
# Remove the intercept row
fixed_effects_no_intercept <- fixed_effects[rownames(fixed_effects) != "(Intercept)", ]

# Reorder the predictors based on the Estimate
fixed_effects_no_intercept$Predictor <- factor(
  rownames(fixed_effects_no_intercept),
  levels = rownames(fixed_effects_no_intercept)[order(fixed_effects_no_intercept$Estimate)]
)

# Plot the fixed effects excluding the intercept, sorted by Estimate
ggplot(fixed_effects_no_intercept, aes(x = Predictor, y = Estimate)) +
  geom_bar(stat = "identity") +
  theme(
    axis.text.x = element_text(angle = 60, hjust = 1, size = 10), # Rotate and adjust the labels
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12)
  ) +
  labs(
    title = "Fixed Effects Coefficients Without Outliers",
    x = "Predictors",
    y = "Estimate"
  )
```


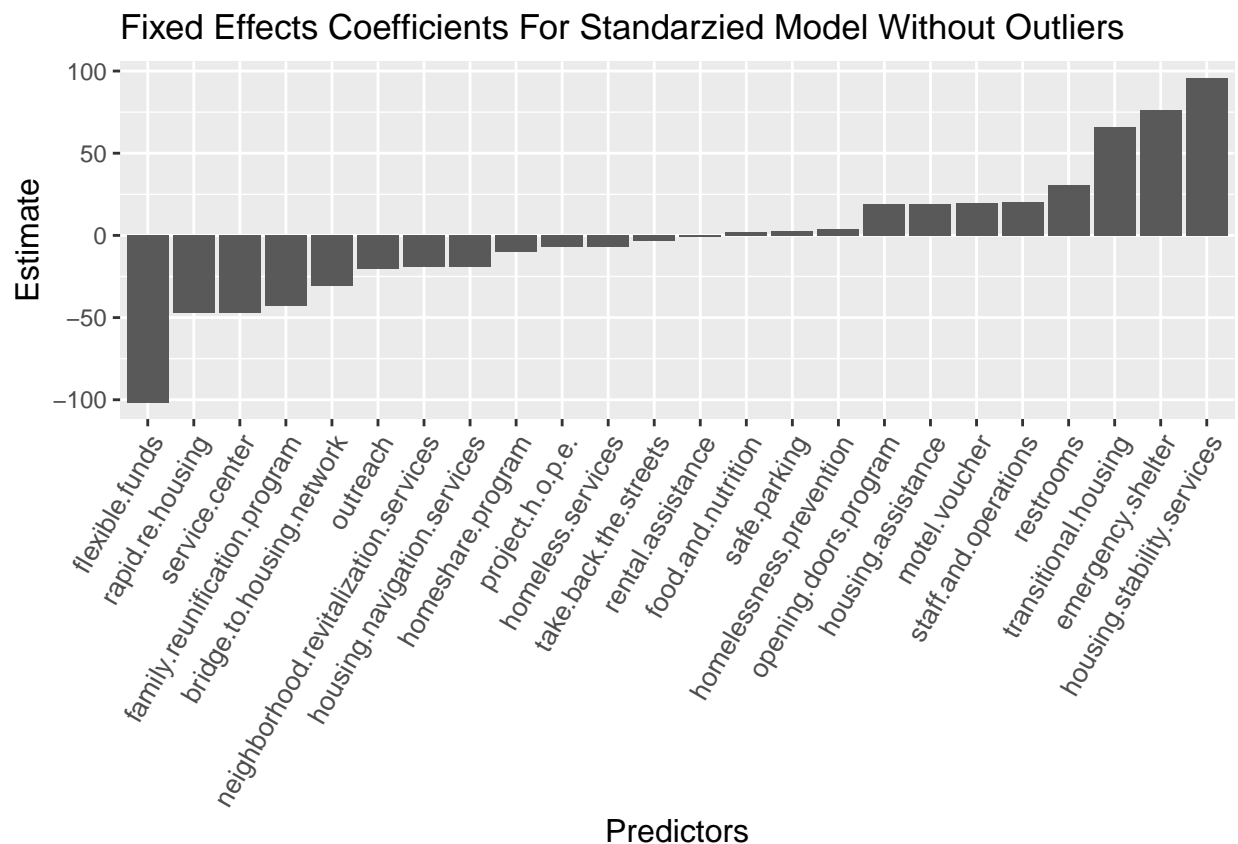Fixed Effects Coefficients Without Outliers

```
# Extract fixed effects
fixed_effectsS <- as.data.frame(summary(model_no_outliersS)$coefficients)

# Remove the intercept row
fixed_effects_no_interceptS <- fixed_effectsS[rownames(fixed_effectsS) != "(Intercept)", ]

# Reorder the predictors based on the Estimate
fixed_effects_no_interceptS$Predictor <- factor(
  rownames(fixed_effects_no_interceptS),
  levels = rownames(fixed_effects_no_interceptS)[order(fixed_effects_no_interceptS$Estimate)]
)

# Plot the fixed effects excluding the intercept, sorted by Estimate
ggplot(fixed_effects_no_interceptS, aes(x = Predictor, y = Estimate)) +
  geom_bar(stat = "identity") +
  theme(
    axis.text.x = element_text(angle = 60, hjust = 1, size = 10), # Rotate and adjust the labels
    axis.title.x = element_text(size = 12),
    axis.title.y = element_text(size = 12)
  ) +
  labs(
    title = "Fixed Effects Coefficients For Standarzied Model Without Outliers",
    x = "Predictors",
    y = "Estimate"
  )
```
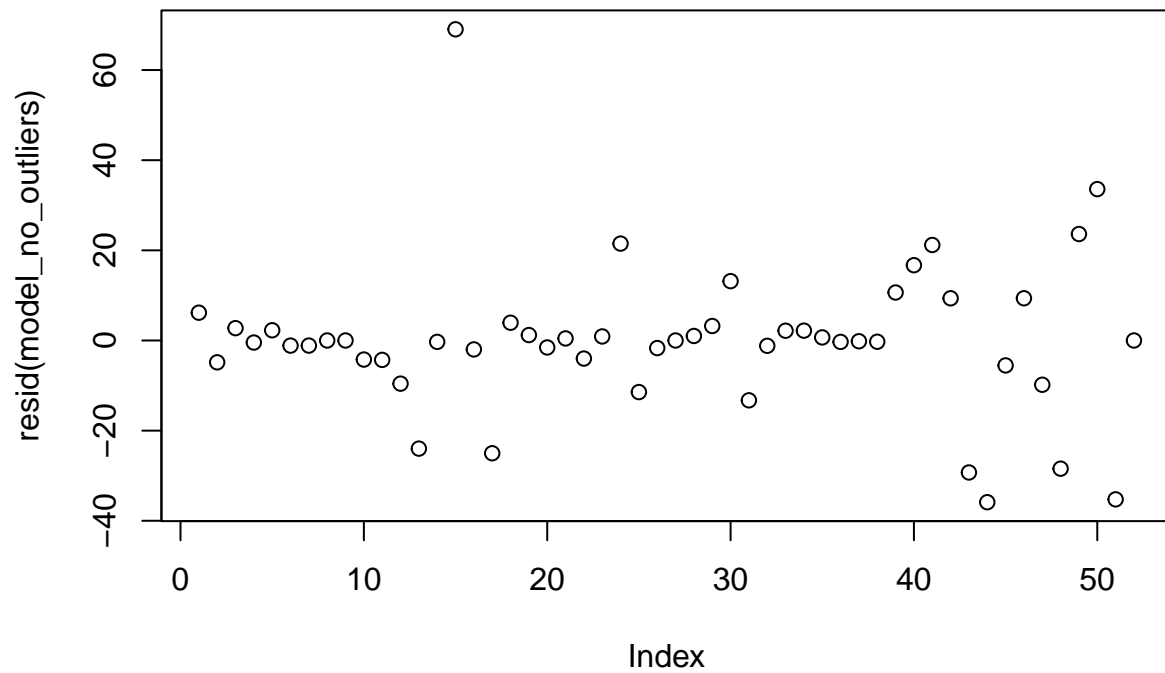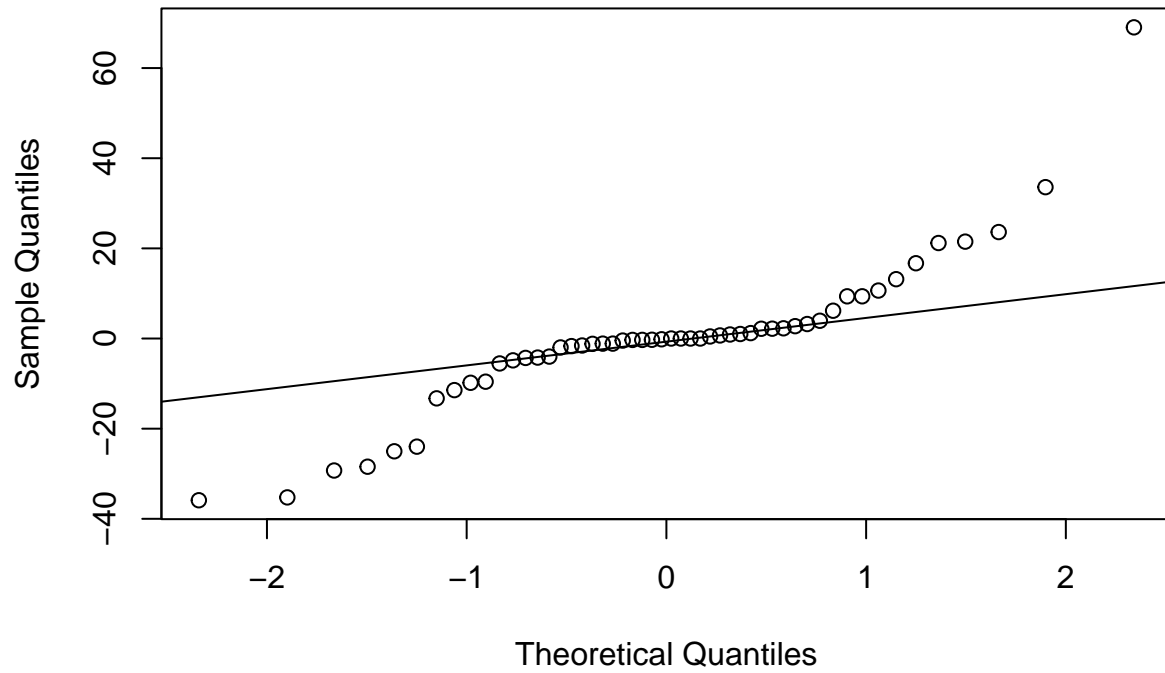

Fixed Effects Coefficients For Standarzied Model Without Outliers

```r
# Residuals and Normality without outliers
plot(resid(model_no_outliers))
```



```r
qqnorm(resid(model_no_outliers))
qqline(resid(model_no_outliers))
```

Normal Q–Q Plot

The order of most to least efficient programs remains consistent even when outliers are removed, only the extent to which they affect Unsheltered PEH per 100,000 is affected, though not by much.