

Model Fitting Process - November 14

Chaeyoung, Hong Gi, Gloria, John, James

December 14, 2022

```
churchData <- read_csv('data/data.csv')
write_csv(churchData, 'data/data.csv')

Church <- read_survey('data/data.csv', legacy = TRUE)

Church <- Church |>
  select(Weeks_Closed, Social_Distancing, Specific_denom,
         Livestreaming, Regions, Signage_of_protocols, Over65_Props) %>%
  na.omit()

summary(Church)
```

##	Weeks_Closed	Social_Distancing	Specific_denom	Livestreaming
##	Min. : 0.00	Mode :logical	Length:701	Mode :logical
##	1st Qu.: 8.00	FALSE:106	Class :character	FALSE:402
##	Median :13.00	TRUE :595	Mode :character	TRUE :299
##	Mean :20.87			
##	3rd Qu.:28.00			
##	Max. :85.00			
##	Regions	Signage_of_protocols	Over65_Props	
##	Length:701	Mode :logical	Min. :1.000	
##	Class :character	FALSE:249	1st Qu.:3.000	
##	Mode :character	TRUE :452	Median :4.000	
##			Mean :4.693	
##			3rd Qu.:7.000	
##			Max. :9.000	

Process Plan

Causal Diagram (copy and paste git hub link below to get image) view at this link: <https://github.com/jmw-75/photo-uploads/raw/main/img245.jpg>

Response Variable: Weeks Closed

Predictors: Specific denomination, Vaccination, Protocols, Region, Small Groups, Social Distancing, livestreaming, In person attendees

Notes: We were considering these predictors but we had some concerns and had to ask our partner for it.
Revisions: We asked our partner about these variables and was told to not include them.

Vaccination is still in consideration since the data were collected after June. Population Question is categorical (figure out a way!)

Our planning process started with evaluating the variables that are available in our data. We took in to our

consideration of how big our data which is 767 rows and 151 columns and consider the possible predictors to use and the ones we have to exclude.

Notes: Another concern that we had to think about but then we talked it out (revision) Region is not an interaction with both, We have to think about if we also want to include region.

Model Fitting

```
model_church <- glmmTMB(Weeks_Closed ~ Social_Distancing
                        + Specific_denom
                        + Livestreaming
                        + Regions * Signage_of_protocols,
                        data = Church,
                        family = nbinom1(link = "log"))

model_church2 <- glmmTMB(Weeks_Closed ~ Social_Distancing
                        + Specific_denom
                        + Livestreaming
                        + Regions * Signage_of_protocols,
                        data = Church,
                        family = nbinom2(link = "log"))

BIC(model_church, model_church2)

##           df      BIC
## model_church  34 5708.837
## model_church2 34 5707.581

summary(model_church)

## Family: nbinom1 ( log )
## Formula:
## Weeks_Closed ~ Social_Distancing + Specific_denom + Livestreaming +
##      Regions * Signage_of_protocols
## Data: Church
##
##      AIC      BIC   logLik deviance df.resid
##  5554.1   5708.8 -2743.0   5486.1      667
##
##
## Dispersion parameter for nbinom1 family (): 12.7
##
## Conditional model:
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.881e+00  4.929e-01   5.846 5.04e-09 ***
## Social_DistancingTRUE -2.136e-01  8.073e-02  -2.646 0.008135 **
## Specific_denomANG    -4.343e-01  4.888e-01  -0.888 0.374281
## Specific_denomB     -3.559e-01  5.057e-01  -0.704 0.481580
## Specific_denomC      1.505e-01  8.937e-01   0.168 0.866275
## Specific_denomCRC     4.211e-02  4.818e-01   0.087 0.930345
## Specific_denomCRC RCA  2.509e-01  6.014e-01   0.417 0.676563
## Specific_denomDOC     8.417e-01  5.525e-01   1.524 0.127623
## Specific_denomECC     6.873e-01  7.779e-01   0.883 0.376983
```

```

## Specific_denomELCA          4.219e-01  4.809e-01   0.877 0.380266
## Specific_denomEVP          -2.894e-01  8.918e-01  -0.324 0.745584
## Specific_denomLMS          -1.663e-01  6.415e-01  -0.259 0.795399
## Specific_denomND           -1.379e-01  4.875e-01  -0.283 0.777324
## Specific_denomORC          -6.100e-01  9.443e-01  -0.646 0.518268
## Specific_denomPCA          -2.486e-01  4.849e-01  -0.513 0.608209
## Specific_denomPCUSA         1.085e+00  6.510e-01   1.667 0.095425
## Specific_denomRCA           8.528e-02  4.836e-01   0.176 0.860036
## Specific_denomSDA          -1.677e+01  2.286e+03  -0.007 0.994146
## Specific_denomUMC          -2.050e-01  7.005e-01  -0.293 0.769749
## Specific_denomWES          -1.265e+00  1.060e+00  -1.193 0.232711
## Specific_denomX             5.138e-01  6.804e-01   0.755 0.450099
## LivestreamingTRUE          4.777e-02  5.690e-02   0.840 0.401142
## RegionsNE                   3.167e-01  1.533e-01   2.066 0.038848 *
## RegionsSE                   -1.523e-01  1.508e-01  -1.010 0.312484
## RegionsSW                   -2.920e-01  2.308e-01  -1.265 0.205987
## RegionsW                     5.736e-01  1.466e-01   3.912 9.14e-05 ***
## RegionsX                    -5.627e-01  5.159e-01  -1.091 0.275378
## Signage_of_protocolsTRUE     4.192e-01  1.082e-01   3.875 0.000107 ***
## RegionsNE:Signage_of_protocolsTRUE -2.896e-01  1.773e-01  -1.633 0.102418
## RegionsSE:Signage_of_protocolsTRUE  1.094e-01  1.705e-01   0.642 0.521052
## RegionsSW:Signage_of_protocolsTRUE -6.931e-03  2.879e-01  -0.024 0.980792
## RegionsW:Signage_of_protocolsTRUE  -5.530e-01  1.708e-01  -3.238 0.001204 **
## RegionsX:Signage_of_protocolsTRUE  -6.612e-02  4.946e-01  -0.134 0.893660
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

We used negative binomial regression model because our response variable is considered as discrete. We ran an AIC test between nbinomial 1 and nbinomial 2 and found that AIC value for nbinomial 1 is lower. Therefore, we will be using the first model (nbinomial 1).

Model Assessment

```

Church <- Church |>
  select(Weeks_Closed, Social_Distancing, Specific_denom,
         Livestreaming, Regions, Signage_of_protocols, Over65_Props) %>%
  na.omit() |>
  mutate(preds = predict(model_church),
         resids = resid(model_church))

```

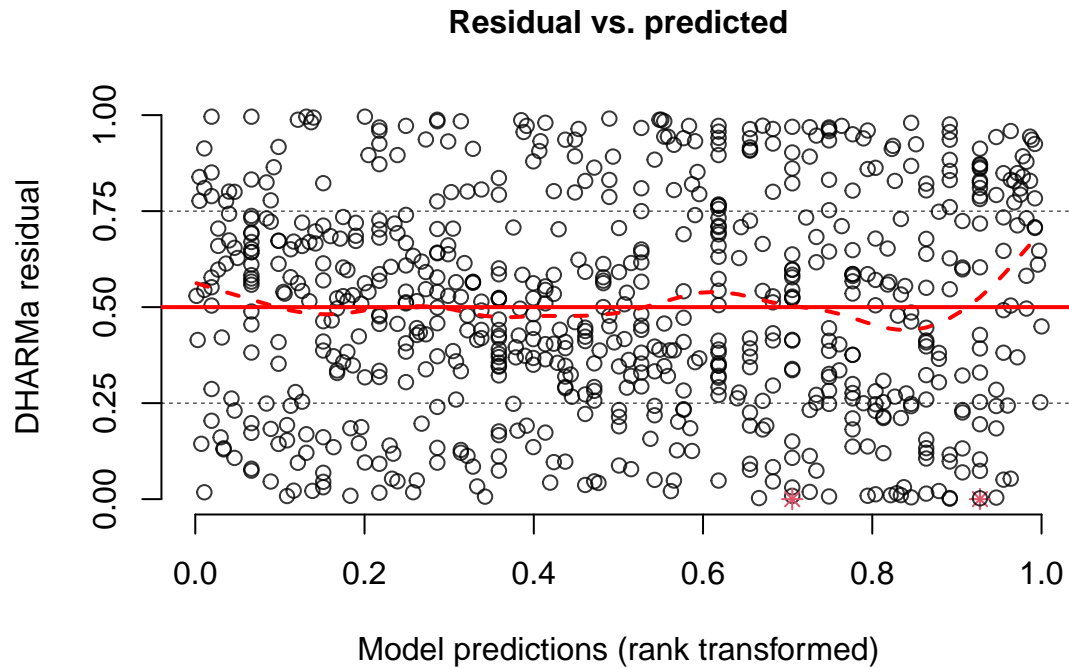
Scaled Residuals

Based on this assessment, we can see that the model passed tests for log linearity and non-constant variance. There does not seem to be a explicit trend in the scaled residuals for linearity. Scaled residuals also appear to be uniform.

```

nb_sim <- simulateResiduals(model_church)
plotResiduals(nb_sim,
              quantreg = FALSE)

```

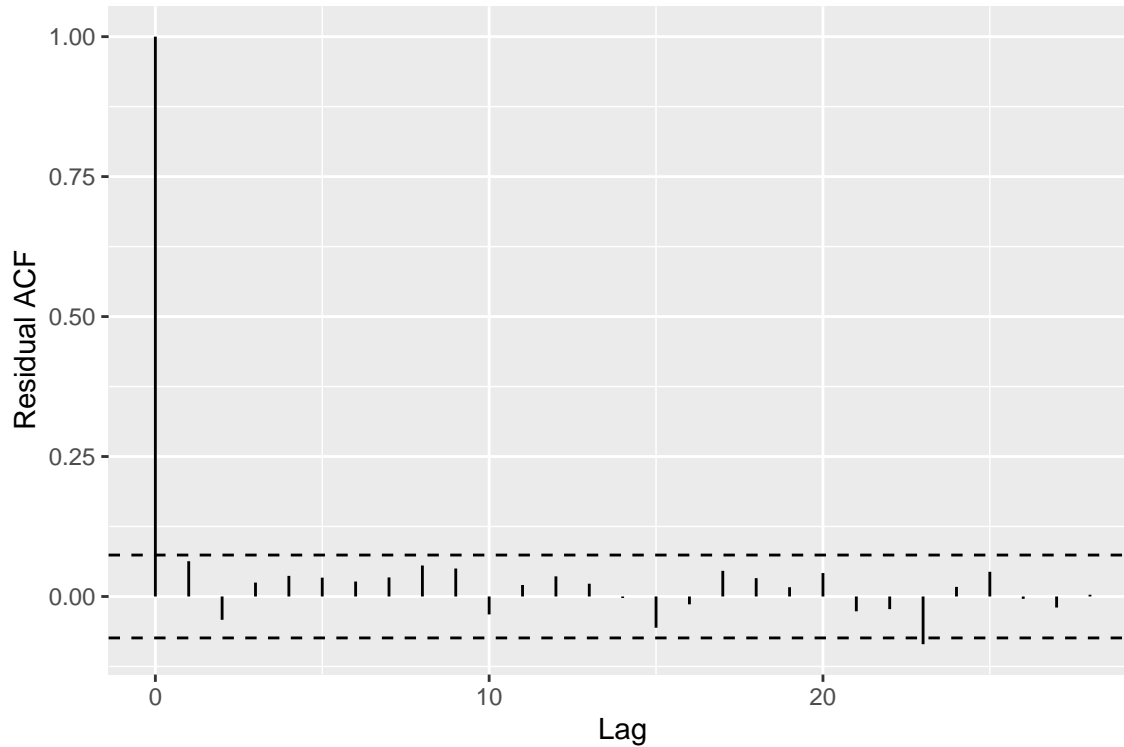


Lack of Non-Linearity

Independence

Model has passed test for independence of variables. No lag spikes are shown in the plot, and all data points are within the bounds set. The data is sorted in order of date recorded. We considered sorting it by the amount of days closed

```
gf_acf(~model_church)
```



Model Selection (if needed)

Because of a large amount of rows, we opted to use BIC. The most optimal BIC value has a model with region and signage of protocols. Along with that, social distancing could be interpreted as a significant predictor, followed by specific denomination. However, when selecting the best model, it would be best to use Region and the Signage of protocols as a interacting effect for it.

```
model_church <- model_church |>
  update(na.action = 'na.fail')
churchDataDredge <- dredge(model_church, rank = 'BIC')
churchDataDredge
```

```
## Global model call: glmmTMB(formula = Weeks_Closed ~ Social_Distancing + Specific_denom +
##   Livestreaming + Regions * Signage_of_protocols, data = Church,
##   family = nbinom1(link = "log"), ziformula = ~0, dispformula = ~1,
##   na.action = "na.fail")
## ---
## Model selection table
##   cnd((Int)) dsp((Int)) cnd(Lvs) cnd(Rgn) cnd(Sgn_of_prt) cnd(Scl_Dst)
## 7         2.872      +          +          +
## 15        2.962      +          +          +          +
## 8         2.869      +          +          +
## 5         2.752      +          +          +
## 16        2.958      +          +          +          +
## 21        2.872      +          +          +
## 29        2.984      +          +          +          +
## 6         2.737      +          +          +
## 13        2.782      +          +          +          +
## 22        2.834      +          +          +
```

## 39	2.773	+		+		+	
## 30	2.944	+	+			+	+
## 14	2.769	+	+			+	+
## 3	3.112	+		+			
## 47	2.887	+		+		+	+
## 17	3.158	+					
## 40	2.772	+	+	+		+	
## 4	3.086	+	+	+			
## 11	3.089	+		+			+
## 23	2.890	+		+		+	
## 18	3.075	+	+				
## 48	2.884	+	+	+		+	+
## 31	3.024	+		+		+	+
## 25	3.193	+					+
## 12	3.067	+	+	+			+
## 24	2.849	+	+	+		+	
## 26	3.115	+	+				+
## 32	2.983	+	+	+		+	+
## 19	3.151	+		+			
## 20	3.067	+	+	+			
## 27	3.223	+		+			+
## 1	3.038	+					
## 55	2.765	+		+		+	
## 28	3.142	+	+	+			+
## 63	2.920	+		+		+	+
## 9	2.910	+					+
## 2	2.994	+	+				
## 10	2.881	+	+				+
## 56	2.728	+	+	+		+	
## 64	2.881	+	+	+		+	+
##	cnd(Spc_dnm)	cnd(Rgn:Sgn_of_prt)	df	logLik	BIC	delta	weight
## 7			8	-2800.324	5653.1	0.00	0.866
## 15			9	-2799.295	5657.6	4.50	0.091
## 8			9	-2800.314	5659.6	6.53	0.033
## 5			3	-2822.029	5663.7	10.65	0.004
## 16			10	-2799.278	5664.1	11.01	0.004
## 21	+		22	-2760.868	5665.9	12.82	0.001
## 29	+		23	-2759.434	5669.6	16.51	0.000
## 6			4	-2821.749	5669.7	16.64	0.000
## 13			4	-2821.911	5670.0	16.96	0.000
## 22	+		23	-2760.548	5671.8	18.74	0.000
## 39			+ 13	-2794.558	5674.3	21.23	0.000
## 30	+		24	-2759.075	5675.4	22.34	0.000
## 14			5	-2821.612	5676.0	22.92	0.000
## 3			7	-2815.130	5676.1	23.06	0.000
## 47			+ 14	-2792.662	5677.1	23.99	0.000
## 17	+		21	-2770.683	5679.0	25.90	0.000
## 40			+ 14	-2794.556	5680.8	27.78	0.000
## 4			8	-2814.759	5681.9	28.87	0.000
## 11			8	-2815.082	5682.6	29.52	0.000
## 23	+		27	-2752.878	5682.7	29.61	0.000
## 18	+		22	-2769.641	5683.4	30.37	0.000
## 48			+ 15	-2792.654	5683.6	30.53	0.000
## 31	+		28	-2750.789	5685.0	31.98	0.000

```
## 25      +      22 -2770.592 5685.3 32.27 0.000
## 12      +      9  -2814.726 5688.4 35.36 0.000
## 24      +      28 -2752.533 5688.5 35.47 0.000
## 26      +      23 -2769.506 5689.7 36.65 0.000
## 32      +      29 -2750.416 5690.9 37.79 0.000
## 19      +      26 -2761.203 5692.8 39.70 0.000
## 20      +      27 -2760.167 5697.3 44.18 0.000
## 27      +      27 -2760.802 5698.5 45.45 0.000
## 1       +      2  -2843.574 5700.3 47.19 0.000
## 55      +      + 32 -2746.589 5702.9 49.79 0.000
## 28      +      28 -2759.696 5702.9 49.79 0.000
## 63      +      + 33 -2743.377 5703.0 49.92 0.000
## 9       +      3  -2842.074 5703.8 50.74 0.000
## 2       +      3  -2842.161 5704.0 50.91 0.000
## 10      +      4  -2840.910 5708.0 54.96 0.000
## 56      +      + 33 -2746.270 5708.8 55.71 0.000
## 64      +      + 34 -2743.026 5708.8 55.77 0.000
## Models ranked by BIC(x)
```

Prediction Plots

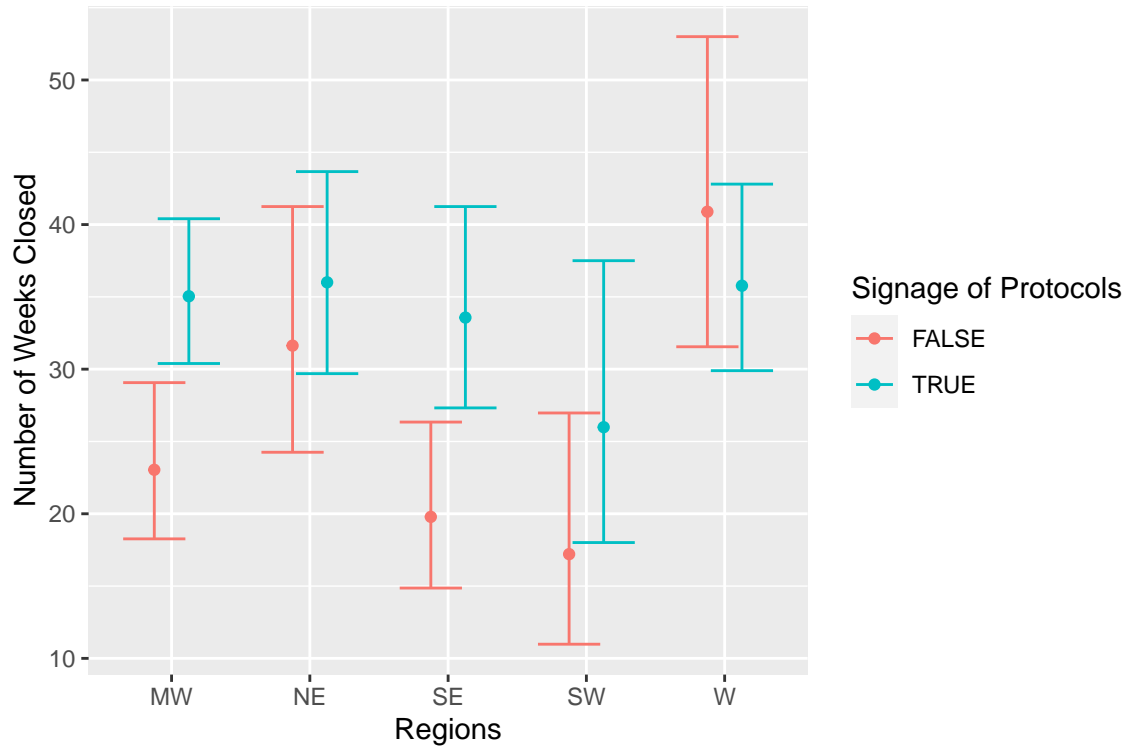
Prediction Plot for Categorical - Categorical Interaction

```
church_pred_data3 <- expand.grid(Specific_denom = 'ELCA',
                                Regions = c('MW', 'NE', 'SE', 'SW', 'W'),
                                Social_Distancing = TRUE,
                                Signage_of_protocols = c(FALSE, TRUE),
                                Livestreaming = TRUE,
                                Over65_Props = 2) |>
  mutate(Signage_of_protocols = factor(as.character(Signage_of_protocols)))
```

```
church_preds <- predict(model_church,
                        newdata = church_pred_data3,
                        type = 'link',
                        se.fit = TRUE)
```

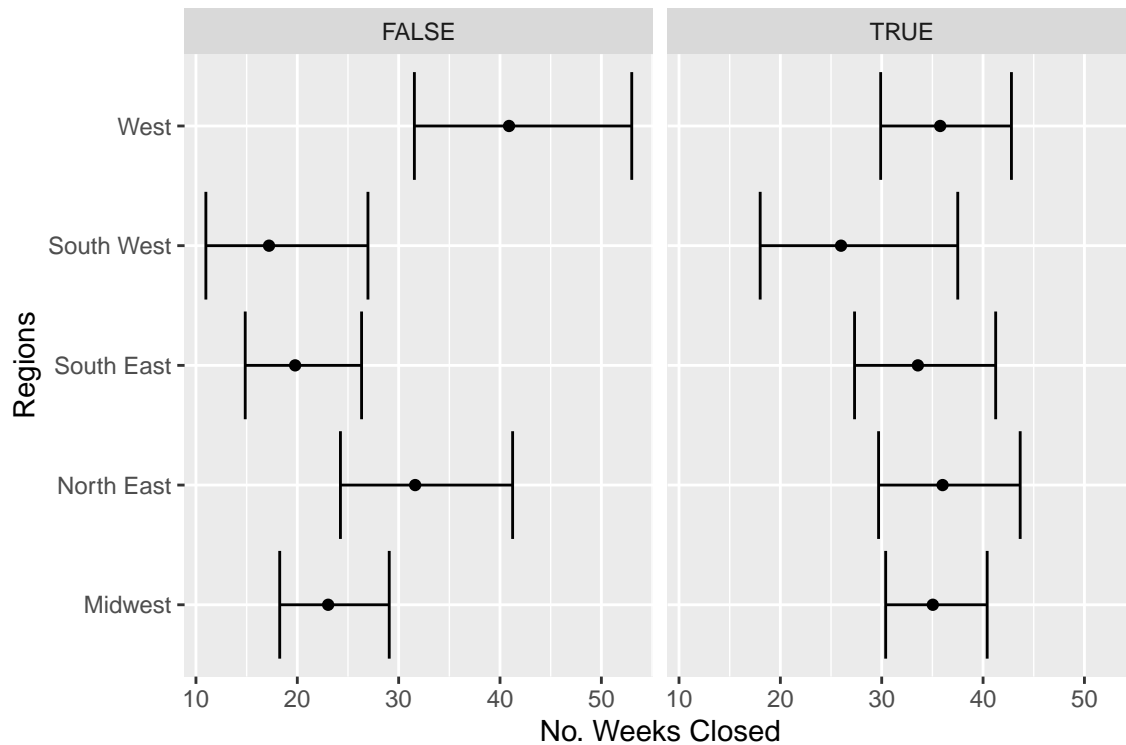
```
church_pred_data3 <- church_pred_data3 |>
  mutate(c_pred = exp(church_preds$fit),
         ci_low = exp(church_preds$fit - 1.96*church_preds$se.fit),
         ci_up = exp(church_preds$fit + 1.96*church_preds$se.fit))
```

```
gf_point(c_pred ~ Regions,
         color = ~Signage_of_protocols,
         position = position_dodge(0.5),
         data = church_pred_data3) |>
  gf_errorbar(ci_low + ci_up ~ Regions,
             color = ~Signage_of_protocols,
             position = position_dodge(0.5)) |>
  gf_labs(y = 'Number of Weeks Closed', color = "Signage of Protocols")
```



Region seems to be significant in predicting weeks closed. This is especially true when determining if the region resided in has COVID protocols signed or not. When unsigned, there is not as much overlap in the regions, indicating higher variance in weeks closed. When signed, one can see that the amount of weeks closed is on average higher than regions where protocols are not signed. This could indicate that as an interacting effect paired with regions is a significant predictor for weeks closed.

```
gf_point(c_pred ~ Regions | Signage_of_protocols, data = church_pred_data3) |>
  gf_errorbar(ci_low + ci_up ~ Regions | Signage_of_protocols)|>
  gf_refine(scale_x_discrete(labels=c("Midwest", "North East", "South East", "South West", "West")), co
  gf_labs(y = 'No. Weeks Closed')
```

Denomination Prediction Plot

```
church_pred_data2 <- expand.grid(Specific_denom = c('PCA', 'CRC', 'ANG', 'B', 'ELCA', 'RCA'),
                                Regions = 'MW',
                                Social_Distancing = TRUE,
                                Signage_of_protocols = TRUE,
                                Livestreaming = TRUE)

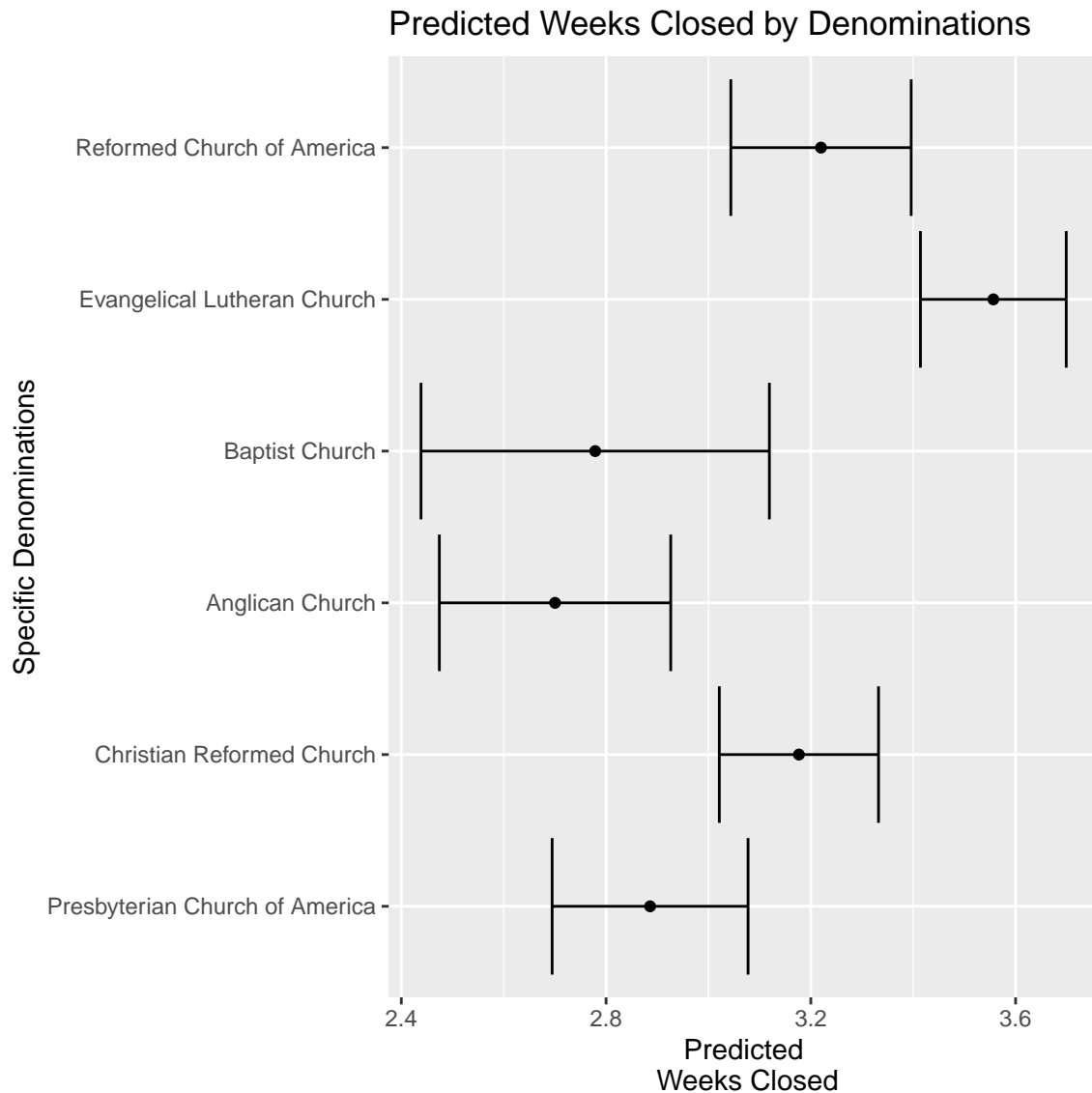
preds <- predict(model_church,
                  newdata = church_pred_data2,
                  se.fit = TRUE,
                  type = "link")
glimpse(preds)

## List of 2
## $ fit : Named num [1:6] 2.89 3.18 2.7 2.78 3.56 ...
## ..- attr(*, "names")= chr [1:6] "eta_predict" "eta_predict" "eta_predict" "eta_predict" ...
## $ se.fit: Named num [1:6] 0.0976 0.0793 0.1153 0.1736 0.0726 ...
## ..- attr(*, "names")= chr [1:6] "eta_predict" "eta_predict" "eta_predict" "eta_predict" ...

church_pred_data2 <- church_pred_data2 |>
  mutate(pred = preds$fit,
         pred.se = preds$se.fit,
         CI_lower = pred - 1.96*pred.se,
         CI_upper = pred + 1.96*pred.se)

gf_point(pred ~ Specific_denom,
         data = church_pred_data2) |>
  gf_labs(y = 'Predicted\n Weeks Closed',
```

```
x = 'Specific Denominations',
title = 'Predicted Weeks Closed by Denominations')|>
gf_refine(scale_x_discrete(labels=c("Presbyterian Church of America", "Christian Reformed Church", "A
gf_errorbar(CI_lower + CI_upper ~ Specific_denom )
```



From this predicted plot, we can see that the denomination of a church can be a significant factor in predicting the amount of weeks closed. Based on denomination of a church, some churches can be predicted to have a higher amount of weeks closed (see Christian reformed and Evangelical), while others may be predicted to have less weeks closed (Baptist, Anglican). Some denominations have some overlap between them, however, due to the nature of how some denomination church culture has some overlap within them as well, denominations could still be used as a significant predictor.

Prediction Plot for Livestreaming

```
church_pred_data4 <- expand.grid(Specific_denom = 'ELCA',
Regions = 'MW',
```

```

      Social_Distancing = TRUE,
      Signage_of_protocols = TRUE,
      Livestreaming = c(TRUE, FALSE)) |>
mutate(Signage_of_protocols = factor(as.character(Signage_of_protocols)))

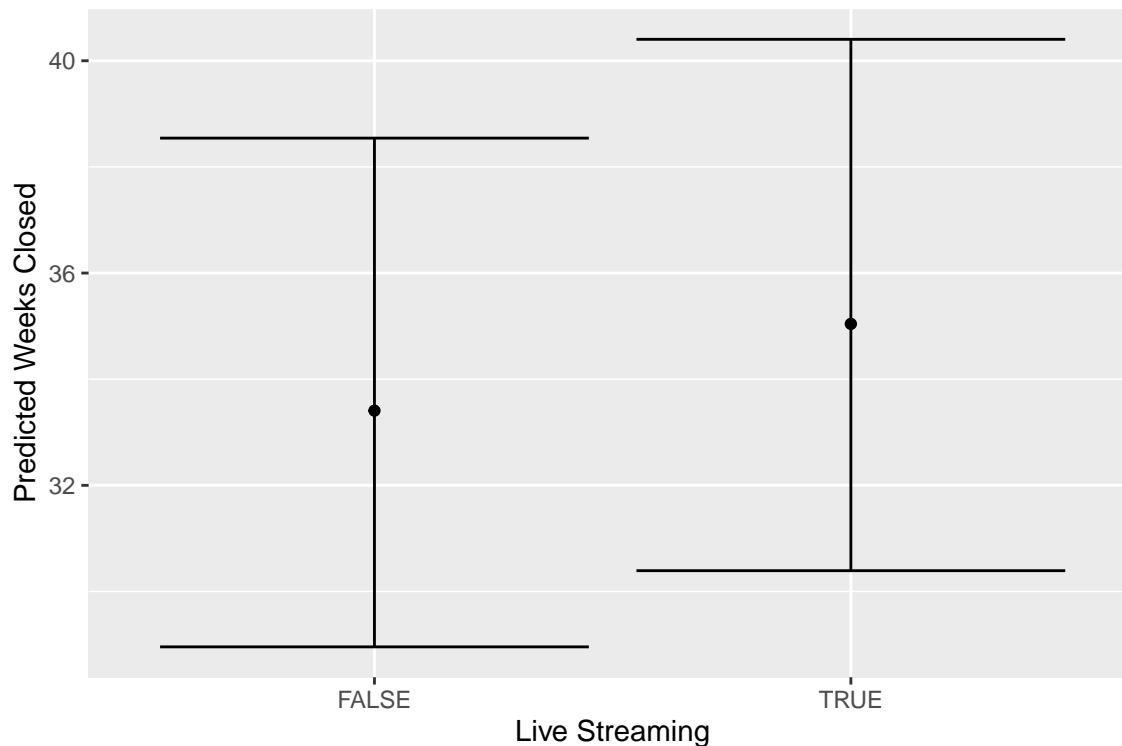
church_predsS <- predict(model_church,
      newdata = church_pred_data4,
      type = 'link',
      se.fit = TRUE)

church_pred_data4 <- church_pred_data4 |>
  mutate(l_pred = church_predsS$fit,
         l_pred.se = church_predsS$se.fit)

church_pred_data4 <- church_pred_data4 |>
  mutate(l_pred = exp(church_predsS$fit),
         ci_low = exp(church_predsS$fit - 1.96*church_predsS$se.fit),
         ci_up = exp(church_predsS$fit + 1.96*church_predsS$se.fit))

gf_point(l_pred ~ Livestreaming, data = church_pred_data4) %>%
  gf_labs(x = "Live Streaming", y = "Predicted Weeks Closed")|>
  gf_errorbar(ci_low + ci_up ~ Livestreaming)

```



Social distancing

```

church_pred_data6 <- expand.grid(Specific_denom = 'ELCA',
      Regions = 'MW',
      Social_Distancing = c(TRUE, FALSE),
      Signage_of_protocols = TRUE,

```

```

Livestreaming = TRUE)

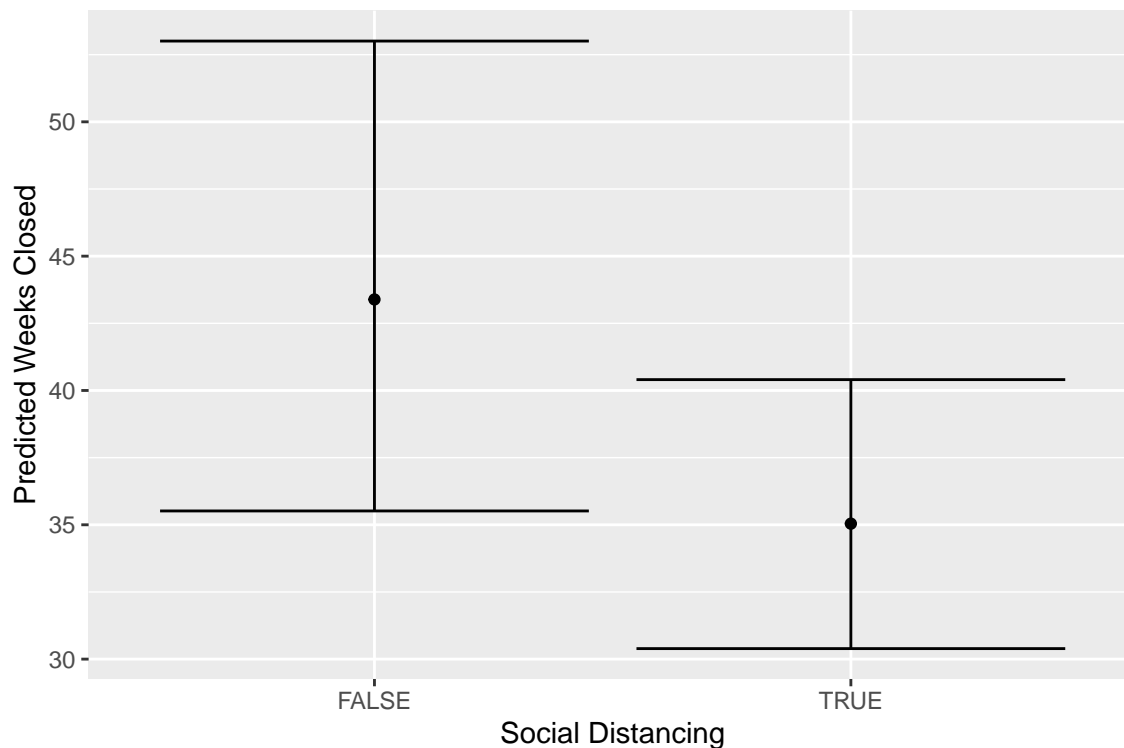
church_preds6 <- predict(model_church,
                          newdata = church_pred_data6,
                          type = 'link',
                          se.fit = TRUE)

church_pred_data6 <- church_pred_data6 |>
  mutate(l_pred = church_preds6$fit,
         l_pred.se = church_preds6$se.fit)

church_pred_data6 <- church_pred_data6 |>
  mutate(l_pred = exp(church_preds6$fit),
         ci_low1 = exp(church_preds6$fit - 1.96*church_preds6$se.fit),
         ci_up1 = exp(church_preds6$fit + 1.96*church_preds6$se.fit))

gf_point(l_pred ~ Social_Distancing, data = church_pred_data6) %>%
  gf_labs(x = "Social Distancing", y = "Predicted Weeks Closed")|>
  gf_errorbar(ci_low1 + ci_up1 ~ Social_Distancing)

```



Over 65 Proportion (asked by our partner)

```

church_pred_data5 <- expand.grid(Specific_denom = 'ELCA',
                                Regions = 'MW',
                                Social_Distancing = TRUE,
                                Signage_of_protocols = TRUE,
                                Livestreaming = TRUE,
                                Over65_Props = seq(from = 1,
                                                    by = 1,
                                                    to = 9))

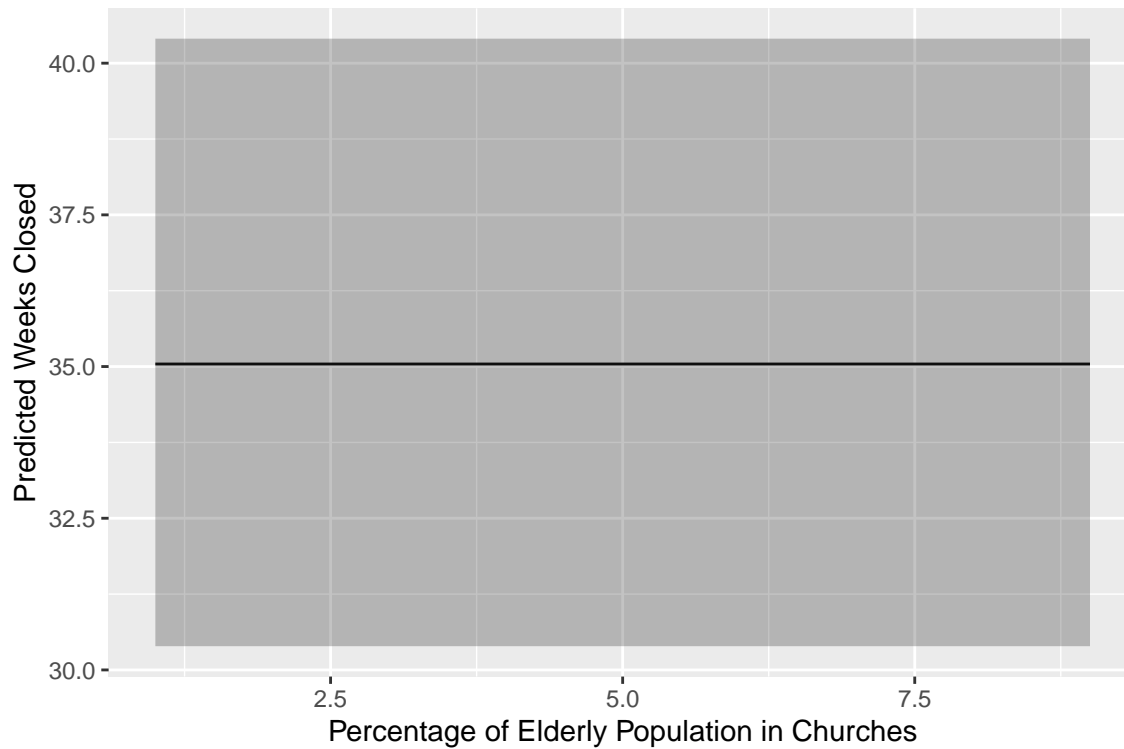
church_preds5 <- predict(model_church,
                        newdata = church_pred_data5,
                        type = 'link',
                        se.fit = TRUE)

church_pred_data5 <- church_pred_data5 |>
  mutate(elderly_pred = church_preds5$fit,
         elderly_pred.se = church_preds5$se.fit)

church_pred_data5 <- church_pred_data5 |>
  mutate(elderly_pred = exp(church_preds5$fit),
         ci_low2 = exp(church_preds5$fit - 1.96*church_preds5$se.fit),
         ci_up2 = exp(church_preds5$fit + 1.96*church_preds5$se.fit))

gf_line(elderly_pred ~ Over65_Props, data = church_pred_data5) %>%
  gf_labs(x = "Percentage of Elderly Population in Churches", y = "Predicted Weeks Closed")|>
  gf_lims(x = c(1,9))|>
  gf_ribbon(ci_low2 + ci_up2 ~ Over65_Props)

```



ELCA

Social Distancing differences between regions of ELCA churches (were asked by our partner to look at)

```
elca_churches <- filter(Church, Specific_denom == "ELCA")

elca_churches %>%
  drop_na(Regions)

## # A tibble: 125 x 9
##   Weeks_Closed Social_Distancing Specific_denom Livestreaming Regions
##   <dbl> <lgl> <chr> <lgl> <chr>
## 1      40 FALSE ELCA TRUE MW
## 2      56 FALSE ELCA FALSE MW
## 3      54 FALSE ELCA FALSE W
## 4      68 TRUE ELCA FALSE W
## 5      23 TRUE ELCA TRUE MW
## 6      10 TRUE ELCA FALSE MW
## 7      11 TRUE ELCA FALSE MW
## 8       1 TRUE ELCA FALSE W
## 9      55 FALSE ELCA TRUE MW
## 10     48 TRUE ELCA FALSE MW
## # ... with 115 more rows, and 4 more variables: Signage_of_protocols <lgl>,
## # Over65_Props <dbl>, preds <dbl>, resid <dbl>
elca_churches <- filter(elca_churches, Regions %in% c("MW", "NE", "SE", "W", "SW"))

gf_bar(~ Regions, fill = ~Social_Distancing, data = elca_churches) |>
  gf_labs(x = "Regions", y = 'Social Distancing')|>
```

```
gf_theme(scale_fill_discrete('Social Distancing?'))
```

