Final Project EDA and Model

Alice Liao, Revo Tesha, Salvador, Cindy, Evelyn

12/1/2019

Exploratory Data Analysis for oh_wv_2012 Dataset

Training data

```
oh_wv_2012 <- read.csv("data/oh_wv_2012.csv", header = TRUE)
set.seed(1000)
train = sample(max(dim(oh_wv_2012)), max(dim(oh_wv_2012))*0.75) # 75% training
train_oh_wv_2012 = oh_wv_2012[train,]
test_oh_wv_2012 = oh_wv_2012[-train,]
```

Distribution of Variables (and possible transformations):

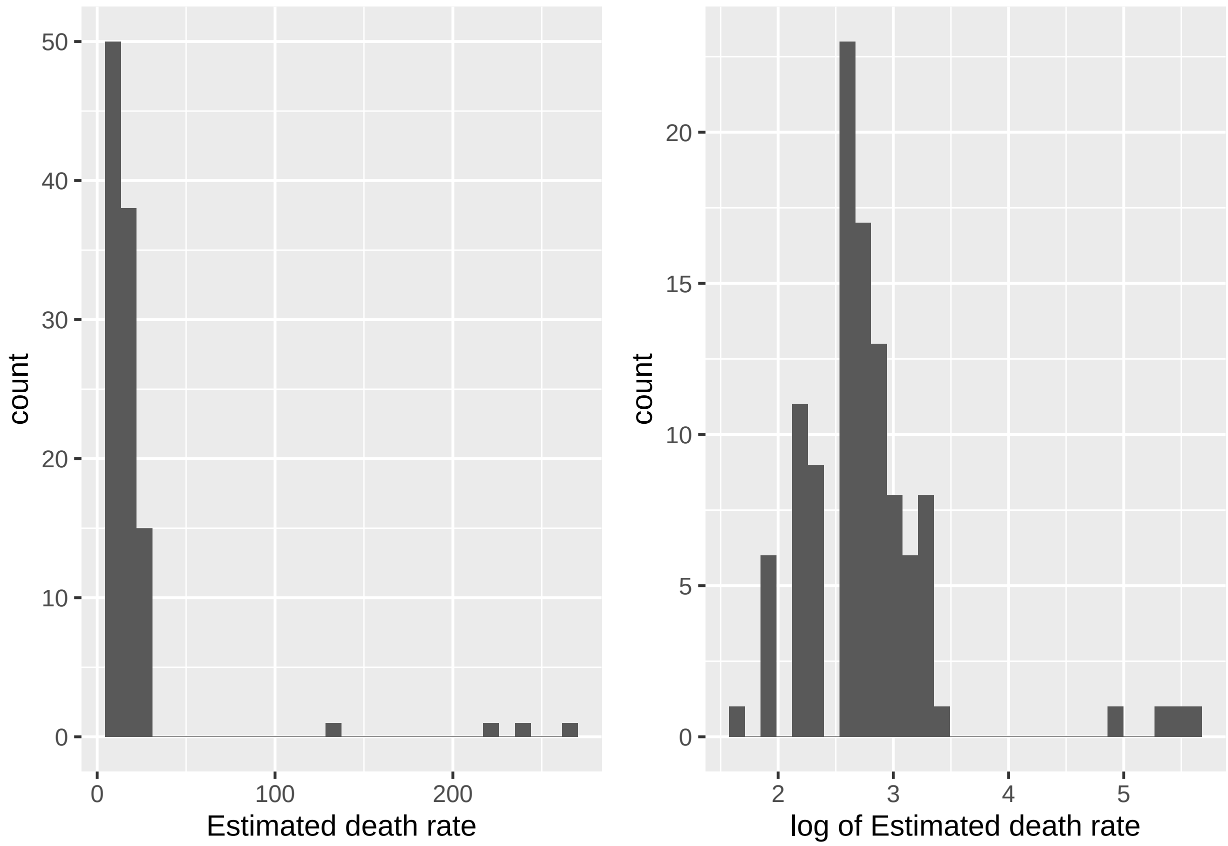
```
summary(train_oh_wv_2012)
```

```
##     BUYER_COUNTY all_active_wt          oxy_wt              hyd_wt
##  ADAMS     :  1   Min.   :   405.7   Min.   :    69.5   Min.   :   336.3
##  ASHLAND   :  1   1st Qu.:  3588.3   1st Qu.:  2072.7   1st Qu.:  1811.0
##  ASHTABULA:  1   Median : 12041.3   Median :  6639.3   Median :  4564.3
##  ATHENS    :  1   Mean   : 30337.6   Mean   : 20650.0   Mean   :  9687.6
##  BARBOUR   :  1   3rd Qu.: 24952.4   3rd Qu.: 17560.8   3rd Qu.:  8765.9
##  BELMONT   :  1   Max.   :448119.7   Max.   :327601.2   Max.   :120518.5
##  (Other)   :101
##     perc_oxy         perc_hyd       perc_retail       perc_chain
##  Min.   :17.10   Min.   :17.90   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:50.85   1st Qu.:30.05   1st Qu.:0.2483   1st Qu.:0.4471
##  Median :61.70   Median :38.30   Median :0.4007   Median :0.5985
##  Mean   :59.06   Mean   :40.94   Mean   :0.4221   Mean   :0.5757
##  3rd Qu.:69.95   3rd Qu.:49.15   3rd Qu.:0.5514   3rd Qu.:0.7512
##  Max.   :82.10   Max.   :82.90   Max.   :1.0000   Max.   :1.0000
##
##  perc_practitioner        most_dist_channel  pharmacy_num
##  Min.   :0.000000   CHAIN PHARMACY :74      Min.   :  1.00
##  1st Qu.:0.000000   RETAIL PHARMACY:33      1st Qu.:  5.00
##  Median :0.001068                           Median : 11.00
##  Mean   :0.002139                           Mean   : 27.29
##  3rd Qu.:0.003272                           3rd Qu.: 25.50
##  Max.   :0.011186                           Max.   :323.00
##
##    distr_num      dominance           State          Year
##  Min.   : 2.00   No :66    Ohio          :66    Min.   :2012
##  1st Qu.: 9.00   Yes:41    West Virginia:41    1st Qu.:2012
##  Median :14.00                                 Median :2012
##  Mean   :15.28                                 Mean   :2012
##  3rd Qu.:20.00                                 3rd Qu.:2012
##  Max.   :49.00                                 Max.   :2012
```

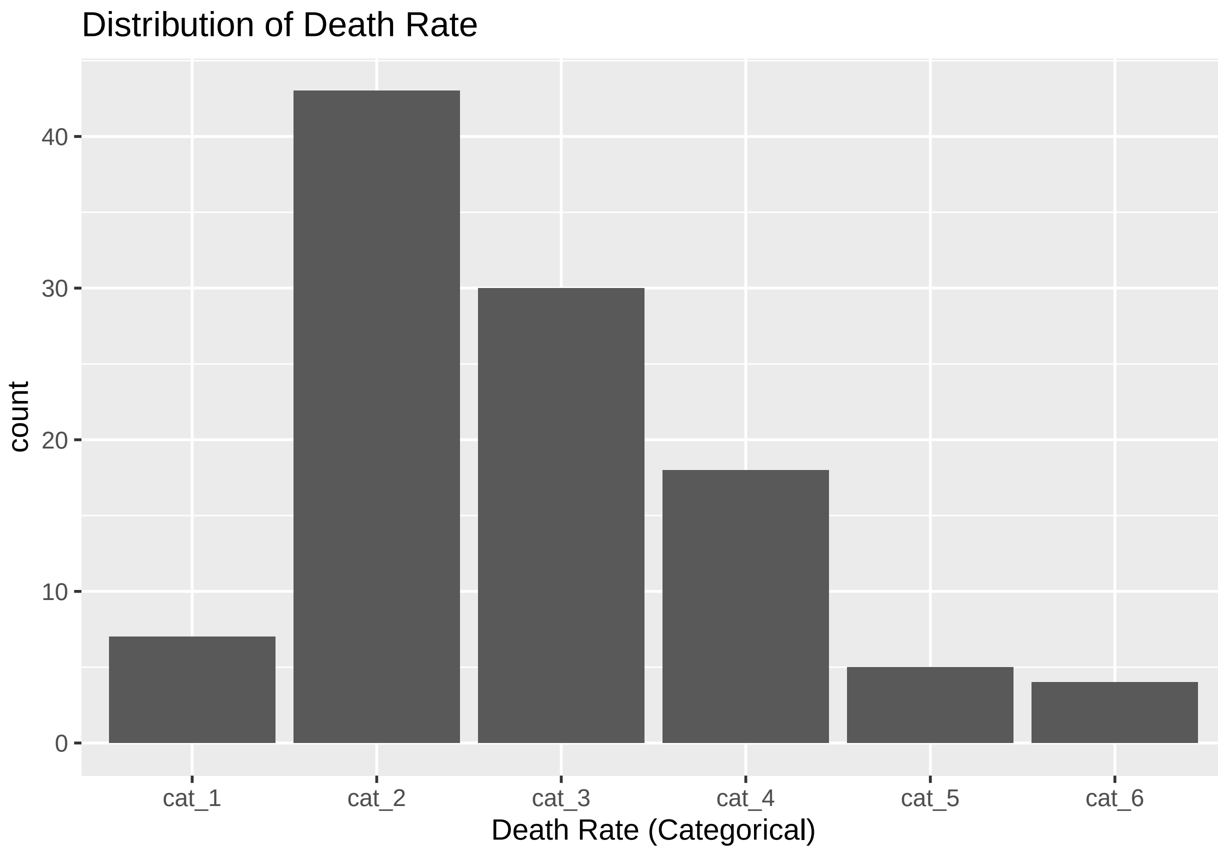
```
##
##  imput_est_death_rate_num est_death_rate_cat   Population
##  Min.   :  4.95           cat_1: 7           Min.   :   5816
##  1st Qu.: 11.95           cat_2:43           1st Qu.:  23640
##  Median : 14.95           cat_3:30           Median :  41856
##  Mean   : 22.67           cat_4:18           Mean   :  99386
##  3rd Qu.: 19.95           cat_5: 5           3rd Qu.:  78256
##  Max.   :262.00           cat_6: 4           Max.   :1265798
##
##  median_income   act_wt_person_county    political_aff pharmacy_num_ptt
##  Min.   :25634   Min.   :0.06512      Democrat  :33    Min.   :0.756
##  1st Qu.:37671   1st Qu.:0.18008      Republican:74    1st Qu.:2.043
##  Median :41877   Median :0.23736                       Median :2.593
##  Mean   :43651   Mean   :0.26721                       Mean   :2.749
##  3rd Qu.:47126   3rd Qu.:0.33498                       3rd Qu.:3.094
##  Max.   :90499   Max.   :0.72302                       Max.   :6.135
##
##  distr_num_ptt
##  Min.   : 0.379
##  1st Qu.: 1.929
##  Median : 2.961
##  Mean   : 3.196
##  3rd Qu.: 4.130
##  Max.   :10.797
##
```

```r
# death rate - numerical
death_rate_p = ggplot(train_oh_wv_2012, aes(x = imput_est_death_rate_num)) +
  geom_histogram() +
  labs(x = "Estimated death rate")
tdeath_rate_p = ggplot(train_oh_wv_2012, aes(x = log(imput_est_death_rate_num))) +
  geom_histogram() +
  labs(x = "log of Estimated death rate")
plot_grid(death_rate_p,tdeath_rate_p)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

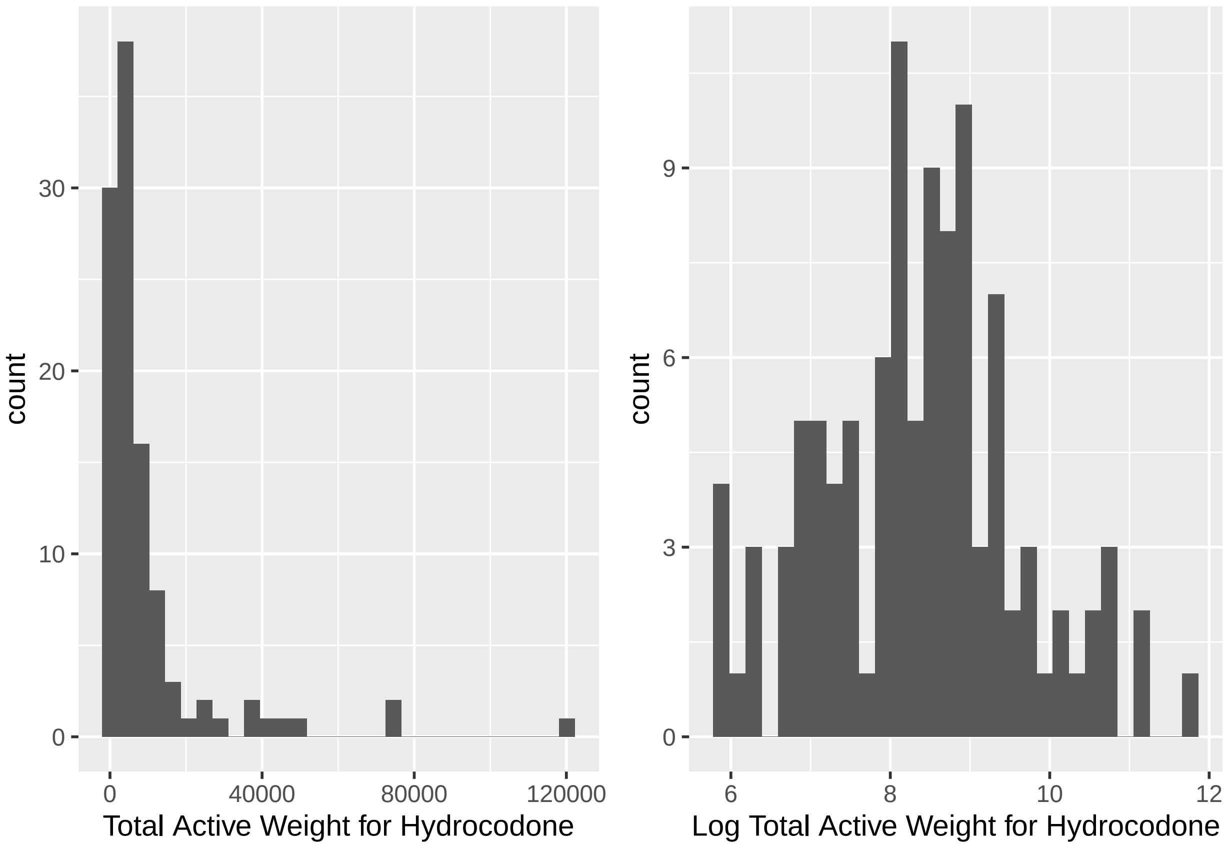



```r
# death rate - categorial
ggplot(train_oh_wv_2012)+
  geom_bar(aes(x = est_death_rate_cat)) +
  labs(x = "Death Rate (Categorical)",
       title = "Distribution of Death Rate")
```



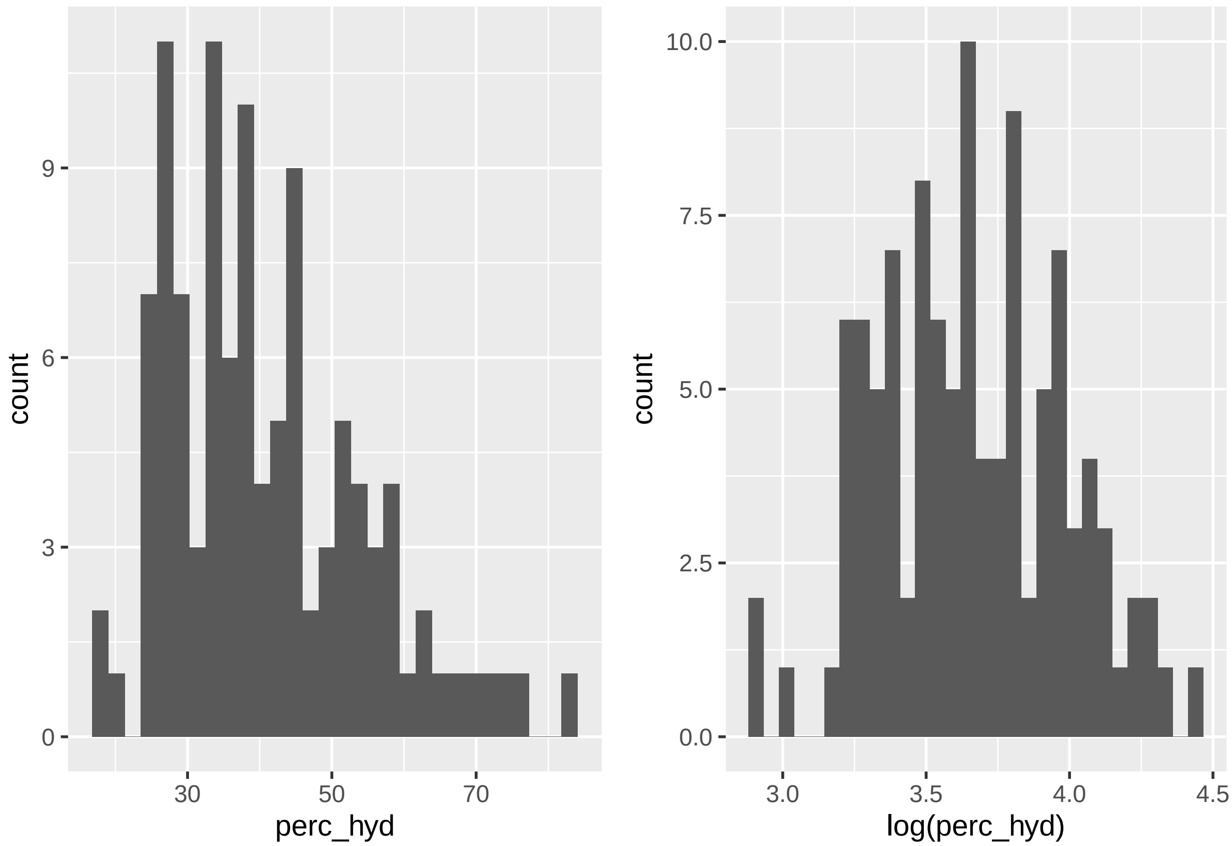

```r
# hydrocodone -- log transformation
hyd_wt_p = ggplot(train_oh_wv_2012, aes(x = hyd_wt)) +
  geom_histogram() +
  labs(x = "Total Active Weight for Hydrocodone")
thyd_wt_p = ggplot(train_oh_wv_2012, aes(x = log(hyd_wt))) +
  geom_histogram() +
  labs(x = "Log Total Active Weight for Hydrocodone")
plot_grid(hyd_wt_p, thyd_wt_p)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



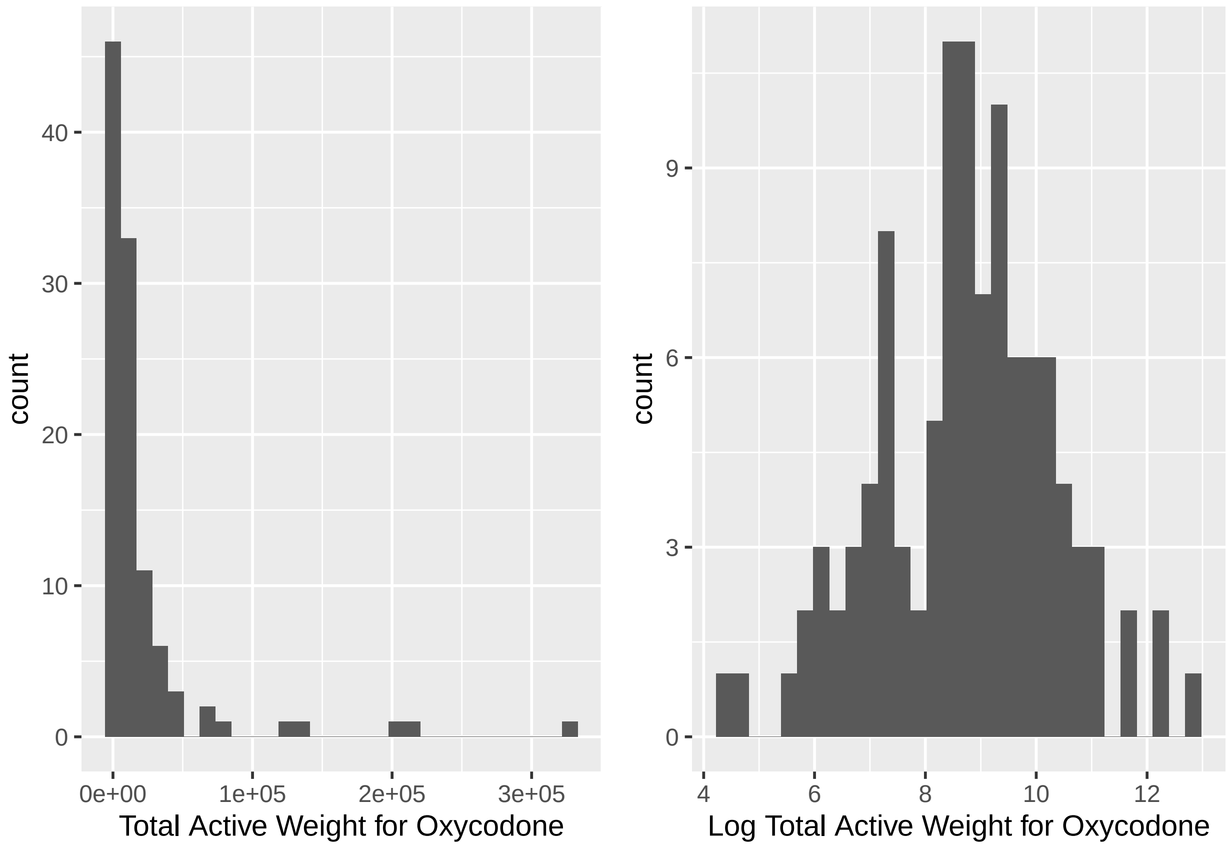

```r
# % hydrocodone
hyd_perc_p = ggplot(train_oh_wv_2012, aes(perc_hyd))+
  geom_histogram()
thyd_perc_p = ggplot(train_oh_wv_2012, aes(log(perc_hyd)))+
  geom_histogram()
plot_grid(hyd_perc_p, thyd_perc_p)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



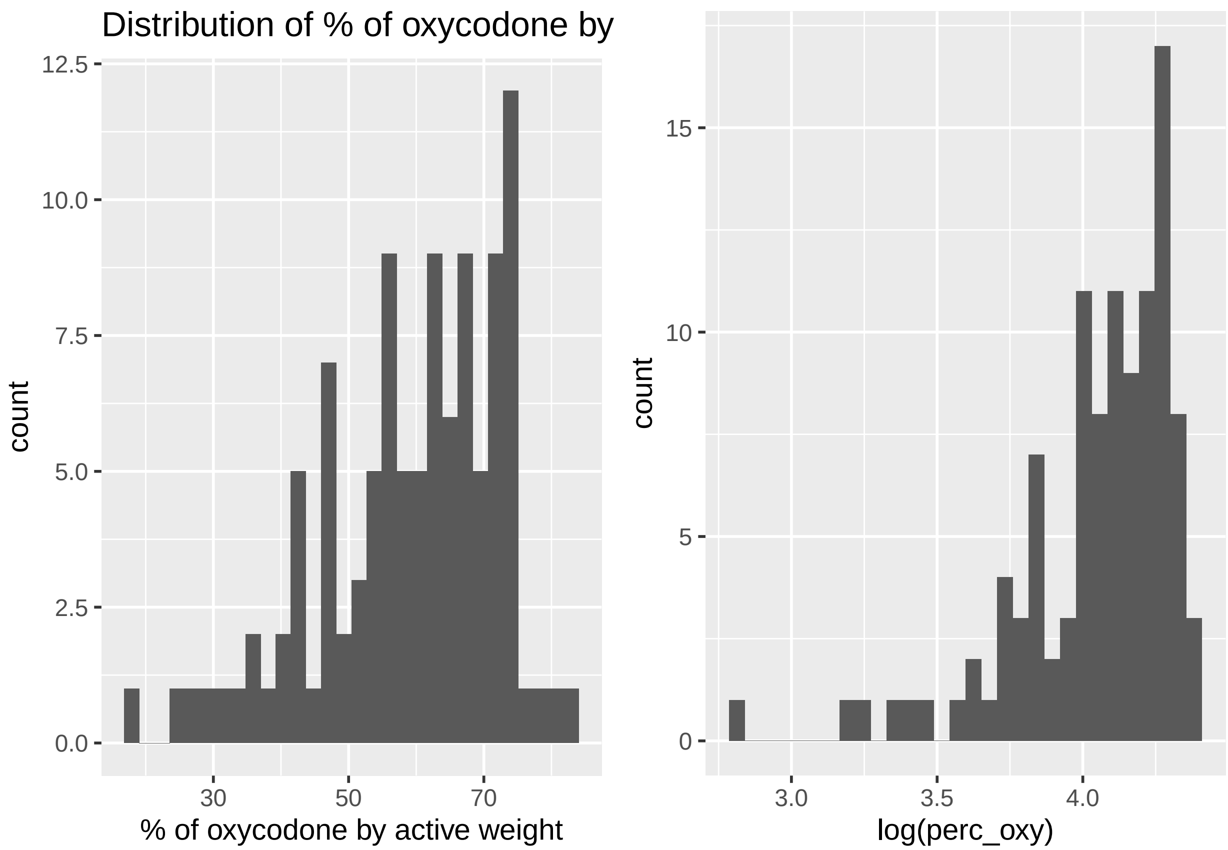

```r
# oxycodone -- log transformation
oxy_wt_p = ggplot(train_oh_wv_2012, aes(x = oxy_wt)) +
  geom_histogram() +
  labs(x = "Total Active Weight for Oxycodone")
toxy_wt_p = ggplot(train_oh_wv_2012, aes(x = log(oxy_wt))) +
  geom_histogram() +
  labs(x = "Log Total Active Weight for Oxycodone")
plot_grid(oxy_wt_p, toxy_wt_p)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



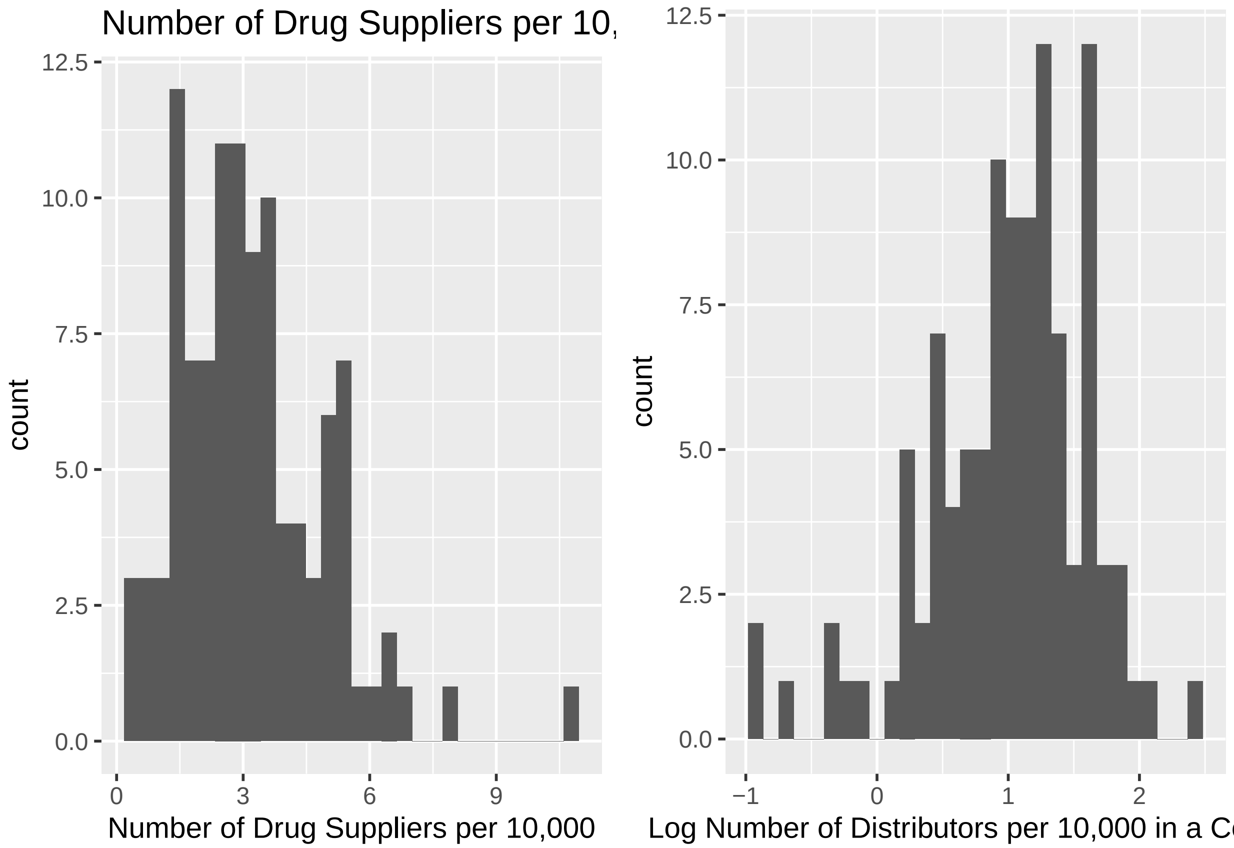

```
# % oxycodone
oxy_perc_p = ggplot(train_oh_wv_2012, aes(perc_oxy))+
  geom_histogram() +
  labs(x = "% of oxycodone by active weight",
       title = "Distribution of % of oxycodone by active weight")

toxy_perc_p = ggplot(train_oh_wv_2012, aes(log(perc_oxy)))+
  geom_histogram()
plot_grid(oxy_perc_p, toxy_perc_p)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



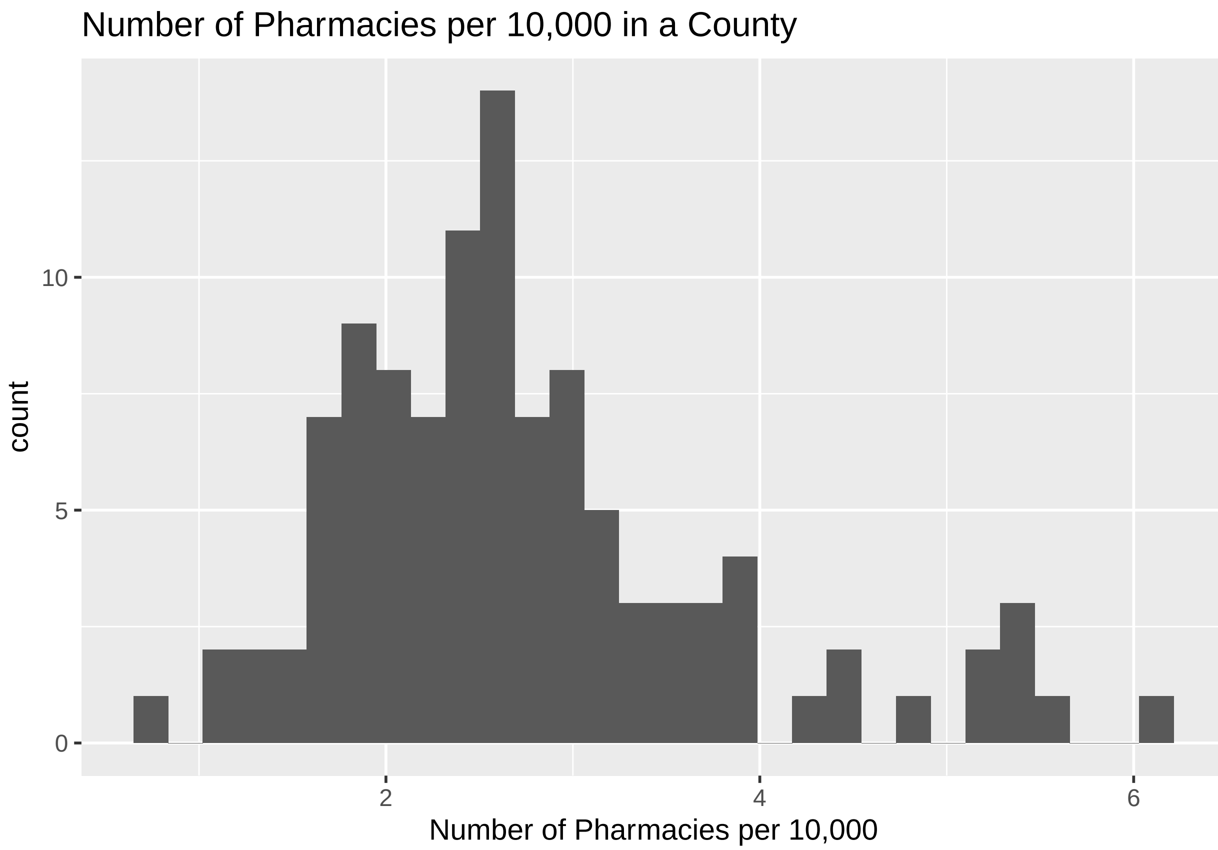

```r
# number of distributors per 10,000 in a county -- no log transformation necessary
ndptt_p = ggplot(train_oh_wv_2012, aes(x = distr_num_ptt)) +
  geom_histogram() +
  labs(x = "Number of Drug Suppliers per 10,000",
       title = "Number of Drug Suppliers per 10,000 in a County")
tndptt_p = ggplot(train_oh_wv_2012, aes(x = log(distr_num_ptt))) +
  geom_histogram() +
  labs(x = "Log Number of Distributors per 10,000 in a County")
plot_grid(ndptt_p, tndptt_p)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



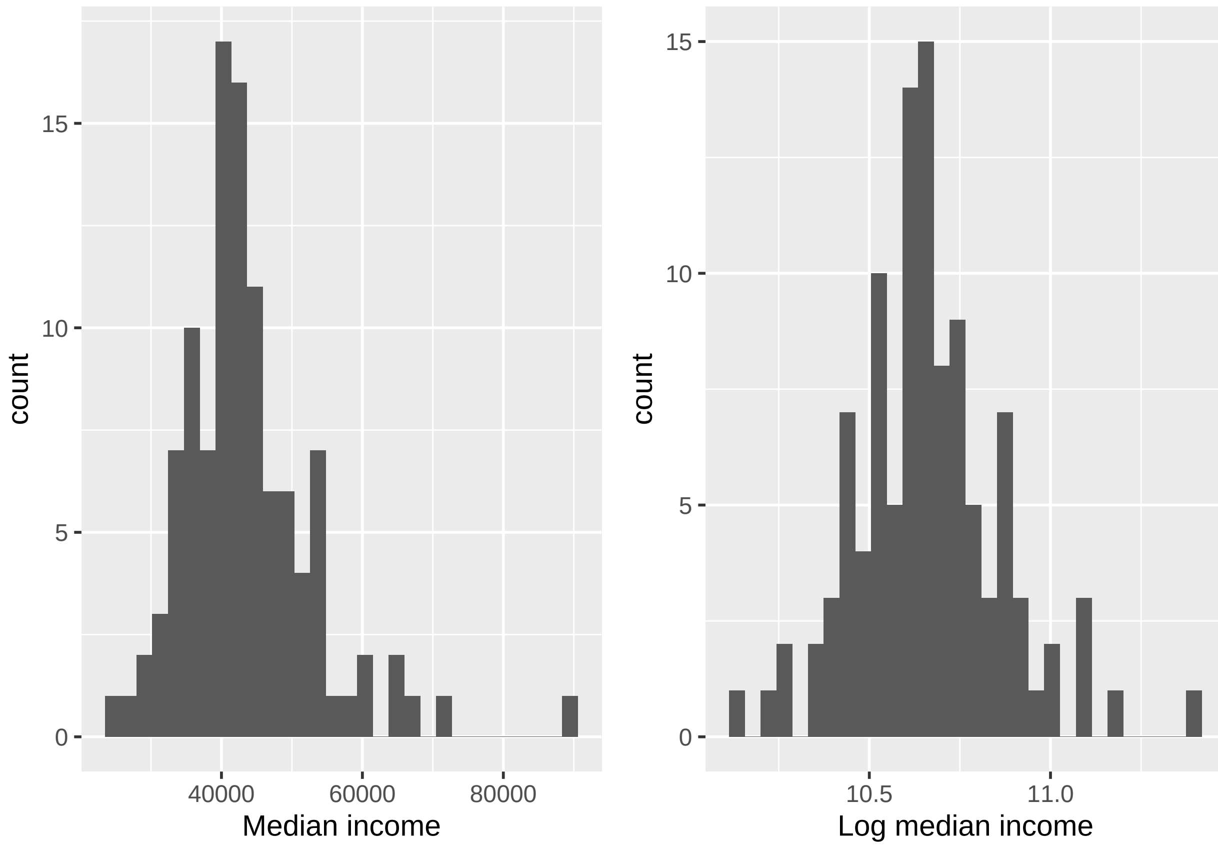

```r
# number of pharmacies per 10,000 for a county -- no transformation necessary
nptt_p = ggplot(train_oh_wv_2012, aes(x = pharmacy_num_ptt)) +
  geom_histogram() +
  labs(x = "Number of Pharmacies per 10,000",
       title = "Number of Pharmacies per 10,000 in a County")
nptt_p
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



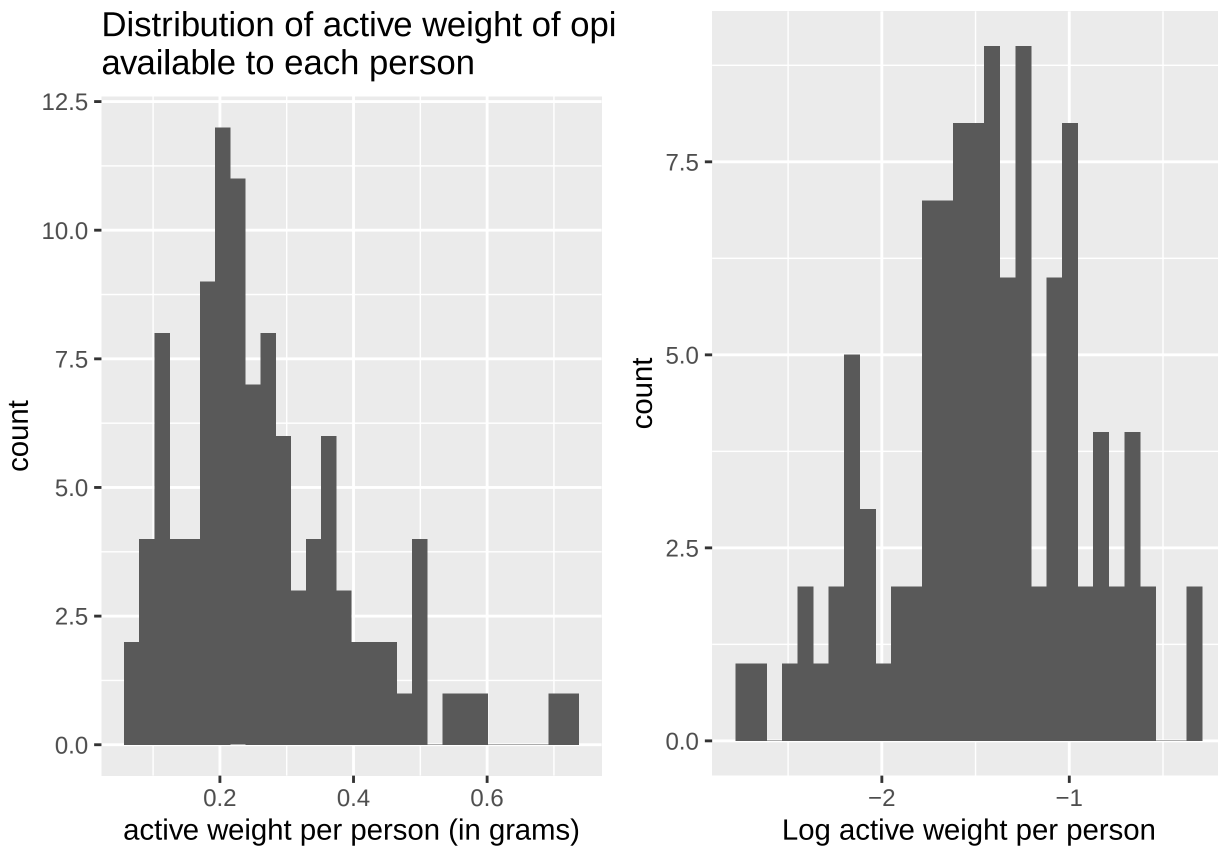

```r
# distribution of median income - need log transformation
med_income = ggplot(train_oh_wv_2012, aes(x = median_income)) +
  geom_histogram() +
  labs(x = "Median income")
log_med_income = ggplot(train_oh_wv_2012, aes(x = log(median_income))) +
  geom_histogram() +
  labs(x = "Log median income")
plot_grid(med_income, log_med_income)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# active drug weight per person in a county -- no log transformation necessary
actwt_pp_plot = ggplot(train_oh_wv_2012, aes(x = act_wt_person_county)) +
  geom_histogram() +
  labs(x = "active weight per person (in grams)",
       title = "Distribution of active weight of opioid \navailable to each person")
log_actwt_pp_plot = ggplot(train_oh_wv_2012, aes(x = log(act_wt_person_county))) +
  geom_histogram() +
  labs(x = "Log active weight per person")
plot_grid(actwt_pp_plot, log_actwt_pp_plot)
```

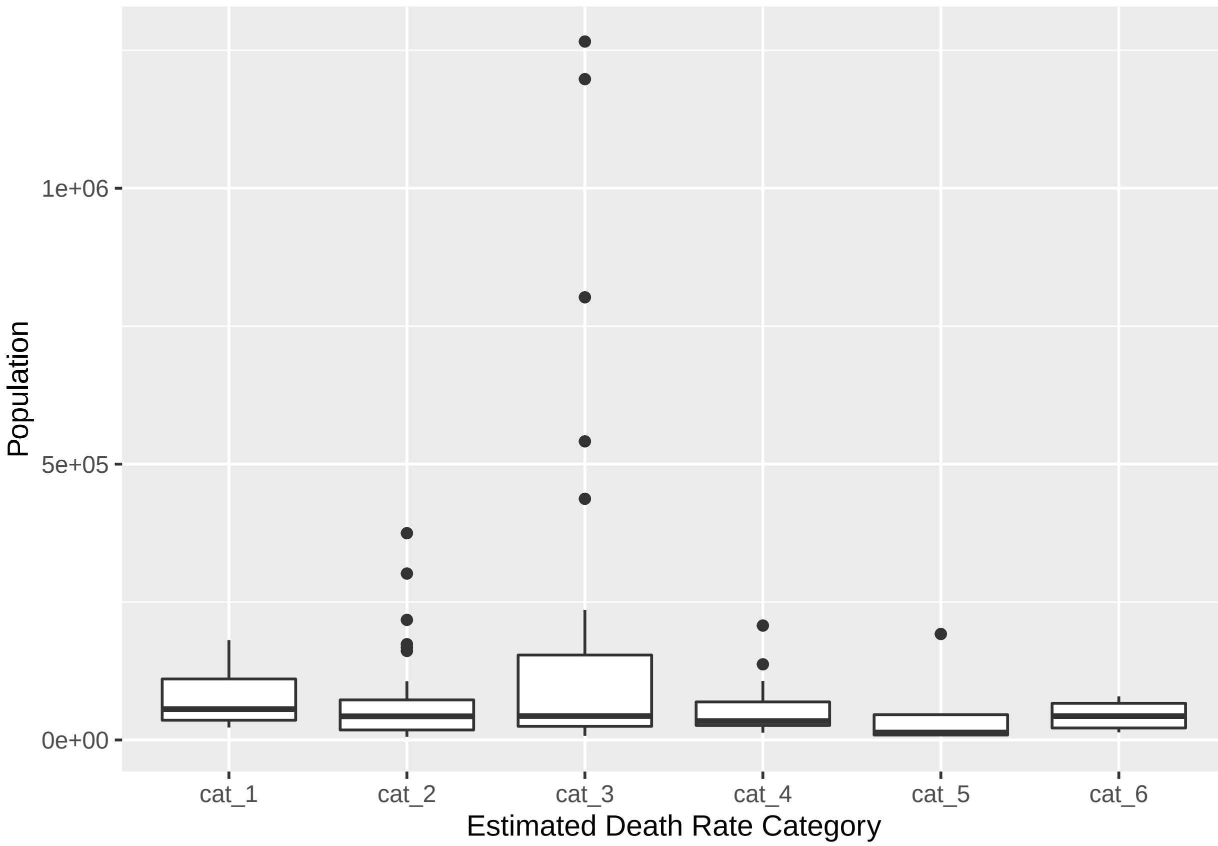
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



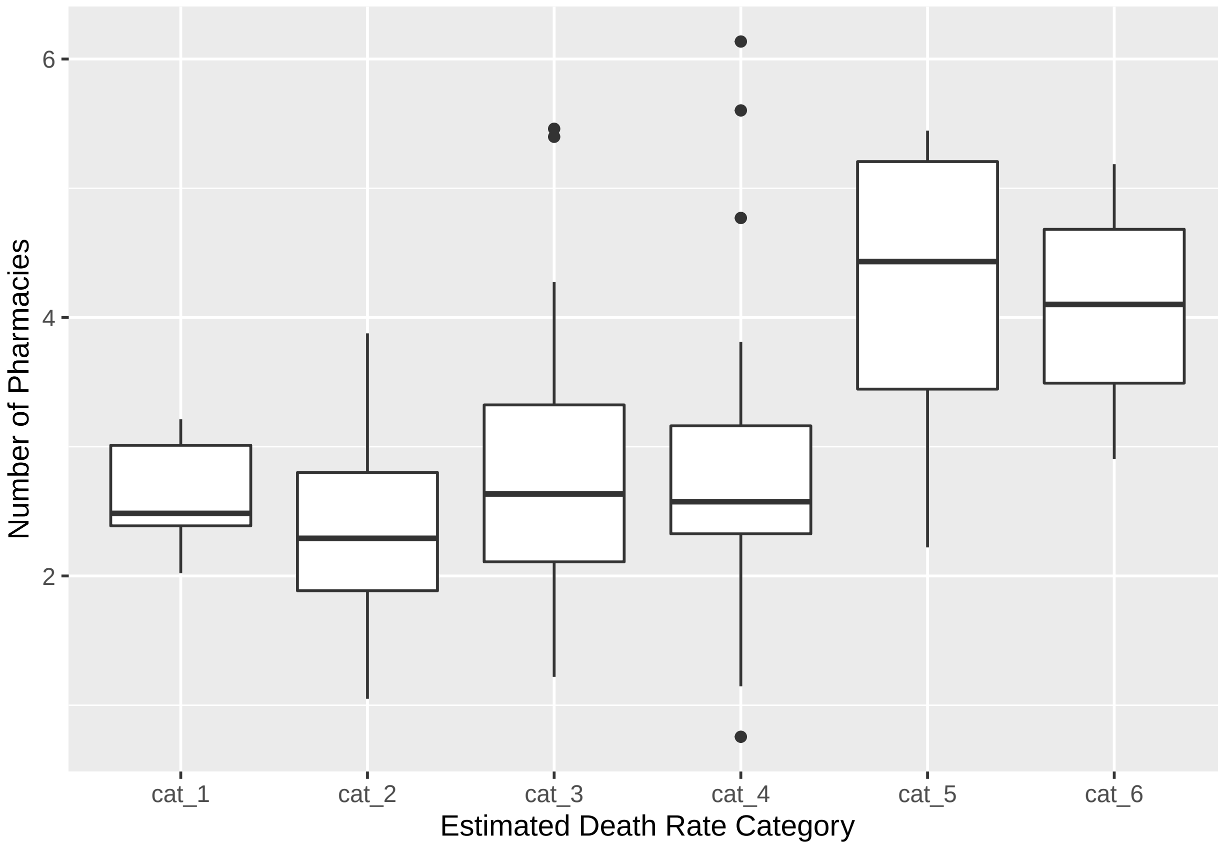
Predictors vs. Response:

Responses: Death Rate Category

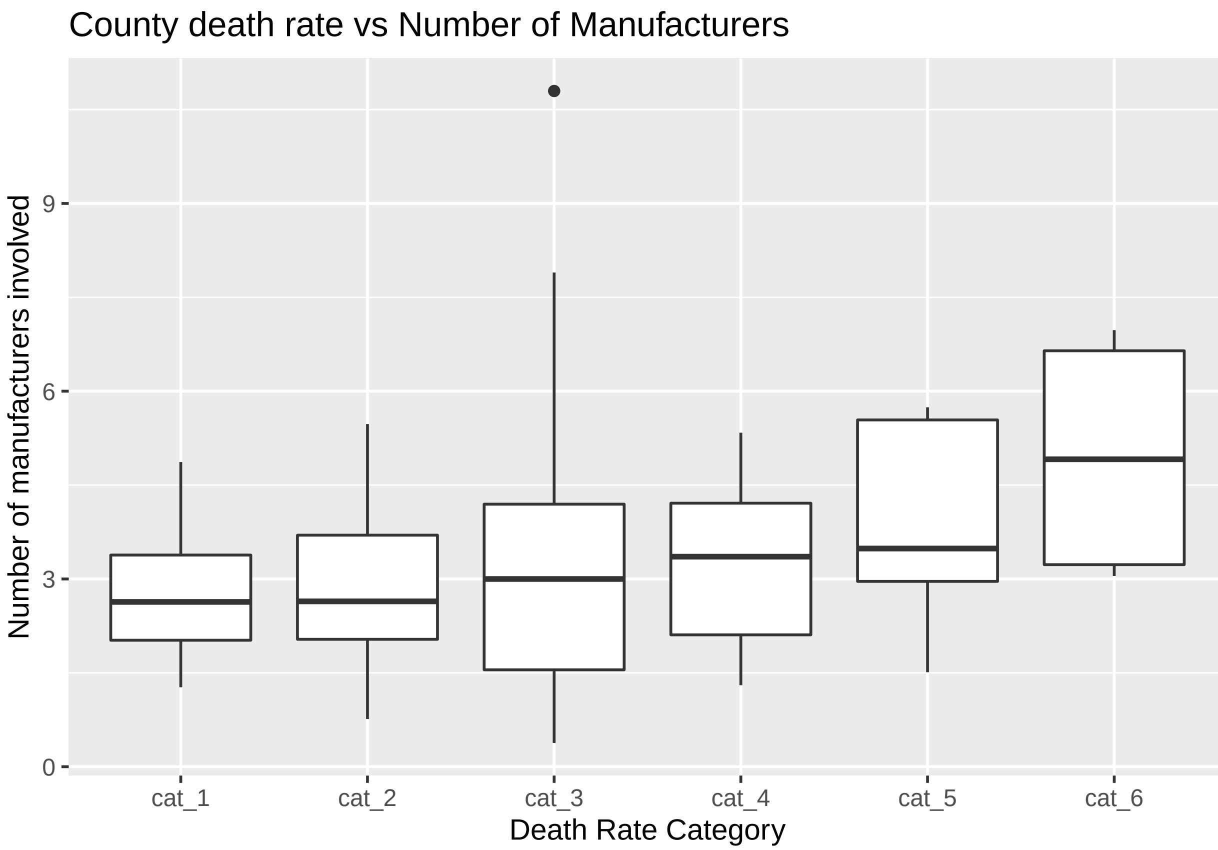
```
ggplot(data = train_oh_wv_2012, aes(x = est_death_rate_cat, y = Population)) +
  geom_boxplot()+
  labs(x = "Estimated Death Rate Category", ylab = "County Population",
        main = "Death Rate Category vs County Population")
```



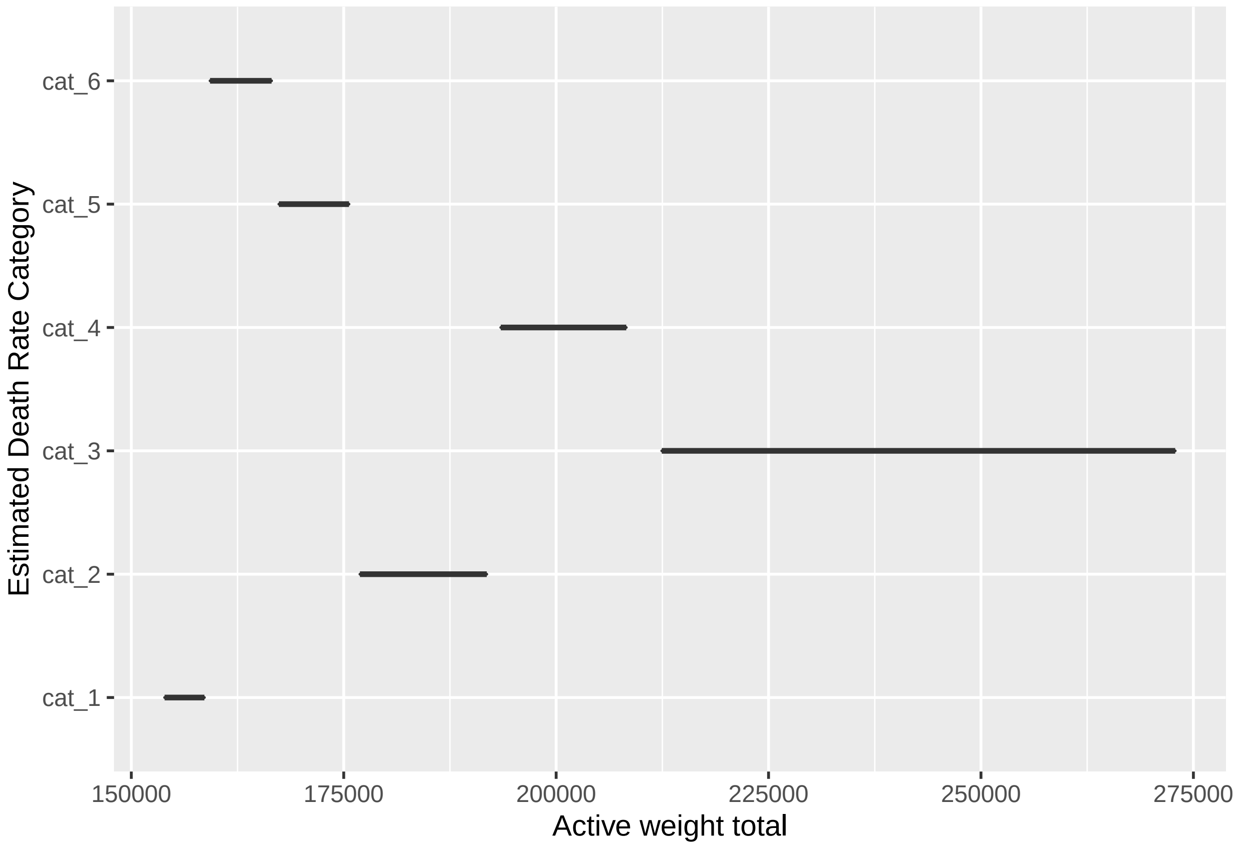

```r
ggplot(data = train_oh_wv_2012, aes(x = est_death_rate_cat, y = pharmacy_num_ptt)) +
  geom_boxplot() +
  labs(x = "Estimated Death Rate Category",
       y = "Number of Pharmacies",
       main = "Death Rate Category vs Number of Pharmacies per 10,000 Population")
```



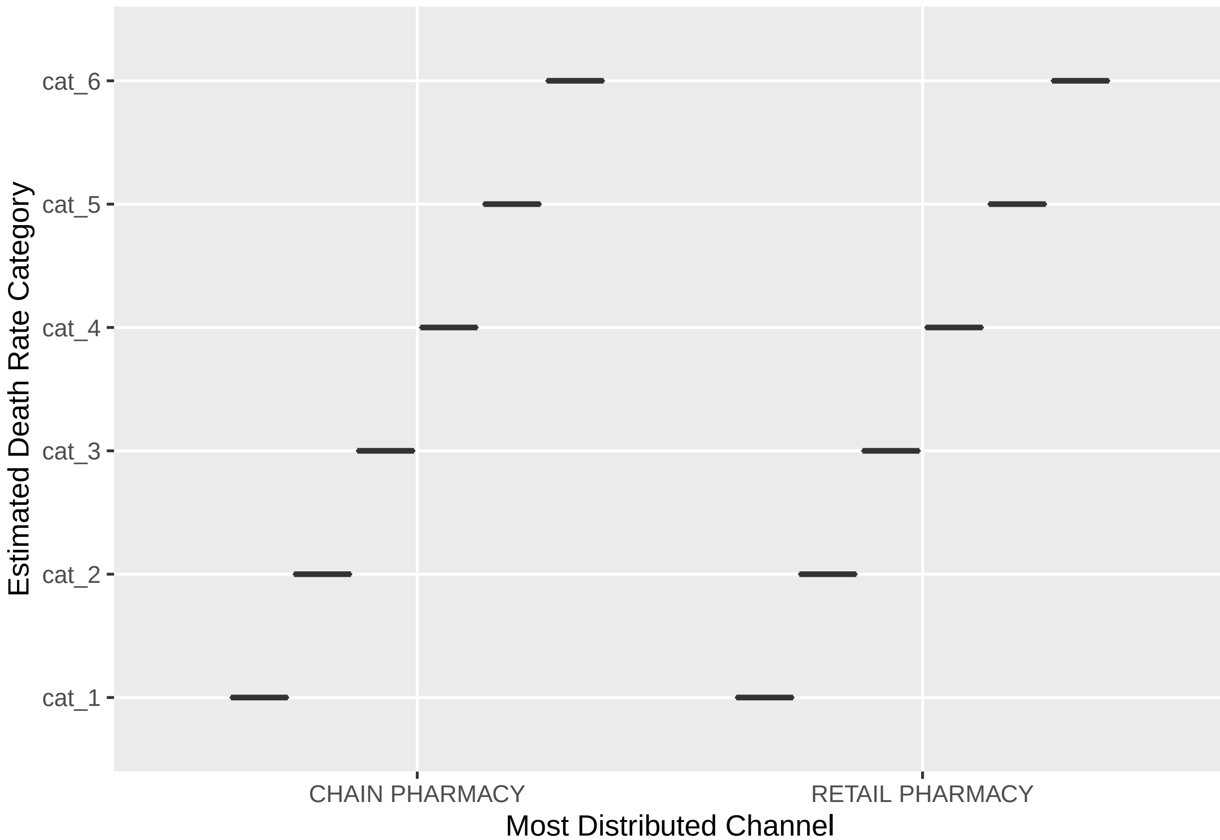

```
ggplot(data = train_oh_wv_2012, aes(x = est_death_rate_cat, y = distr_num_ptt)) +
  geom_boxplot() +
  labs(x = "Death Rate Category", y = "Number of manufacturers involved",
       title = "County death rate vs Number of Manufacturers")
```




```
##################################################################################
ggplot(data = train_oh_wv_2012, aes(x = all_active_wt, y = est_death_rate_cat)) +
  geom_boxplot() + xlab("Active weight total") + ylab("Estimated Death Rate Category")
```

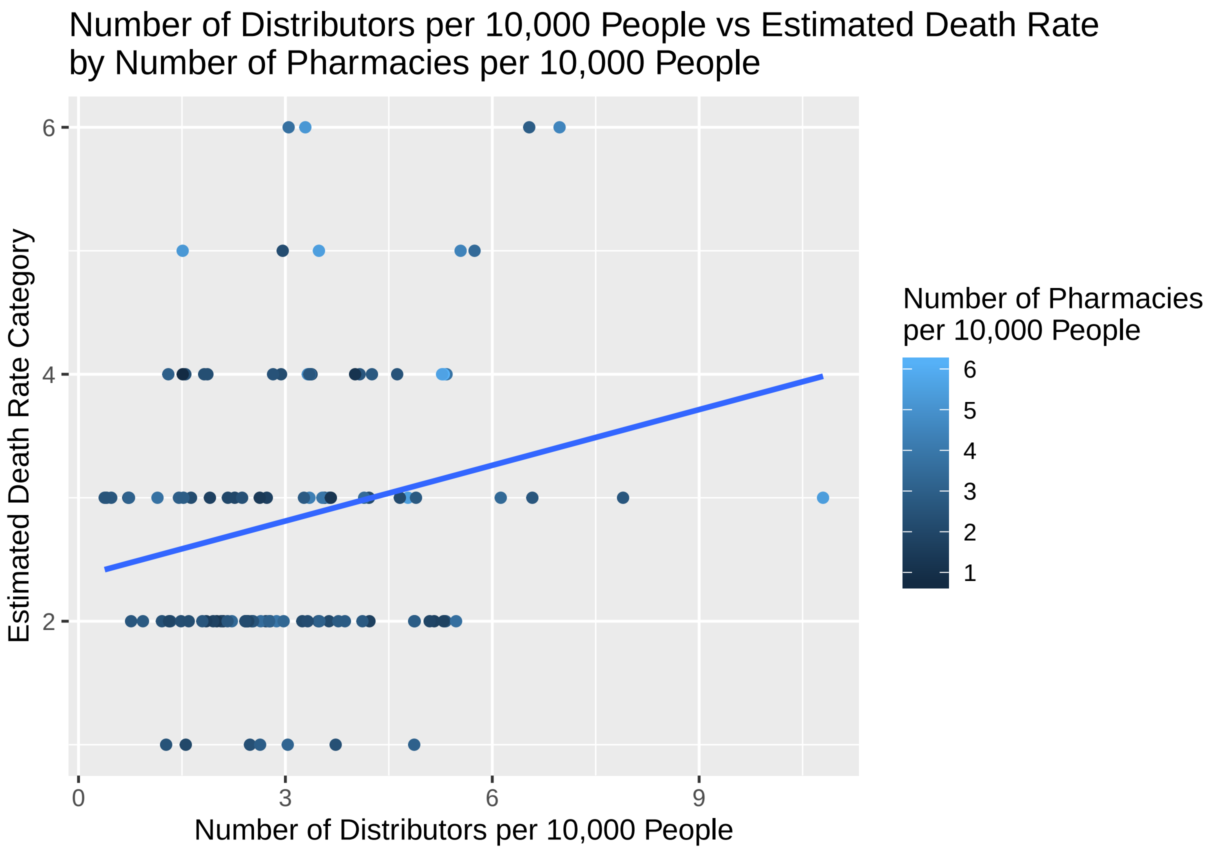



```
train_oh_wv_2012 %>%
  ggplot(aes(x = most_dist_channel, y = est_death_rate_cat)) + geom_boxplot() + xlab("Most Distributed C
```

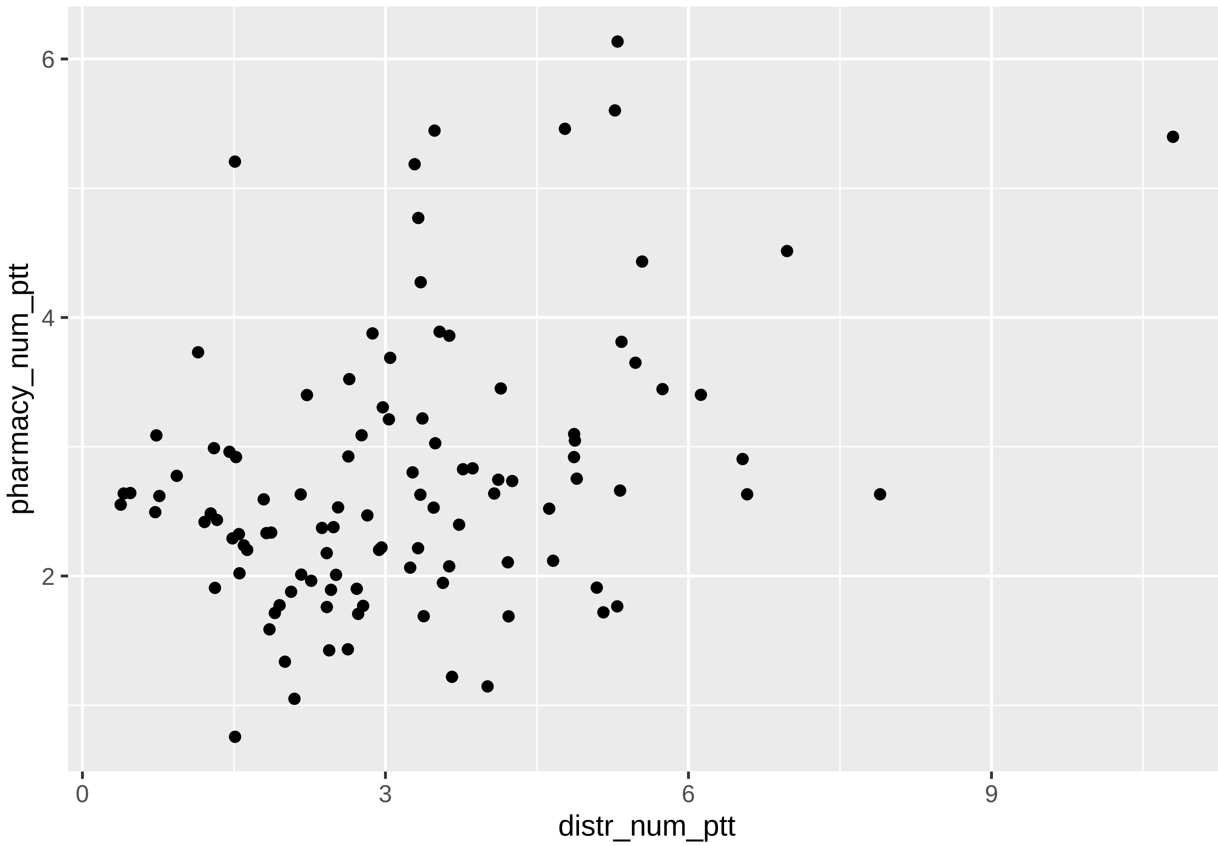


Plot predictors against each other (Interactions)

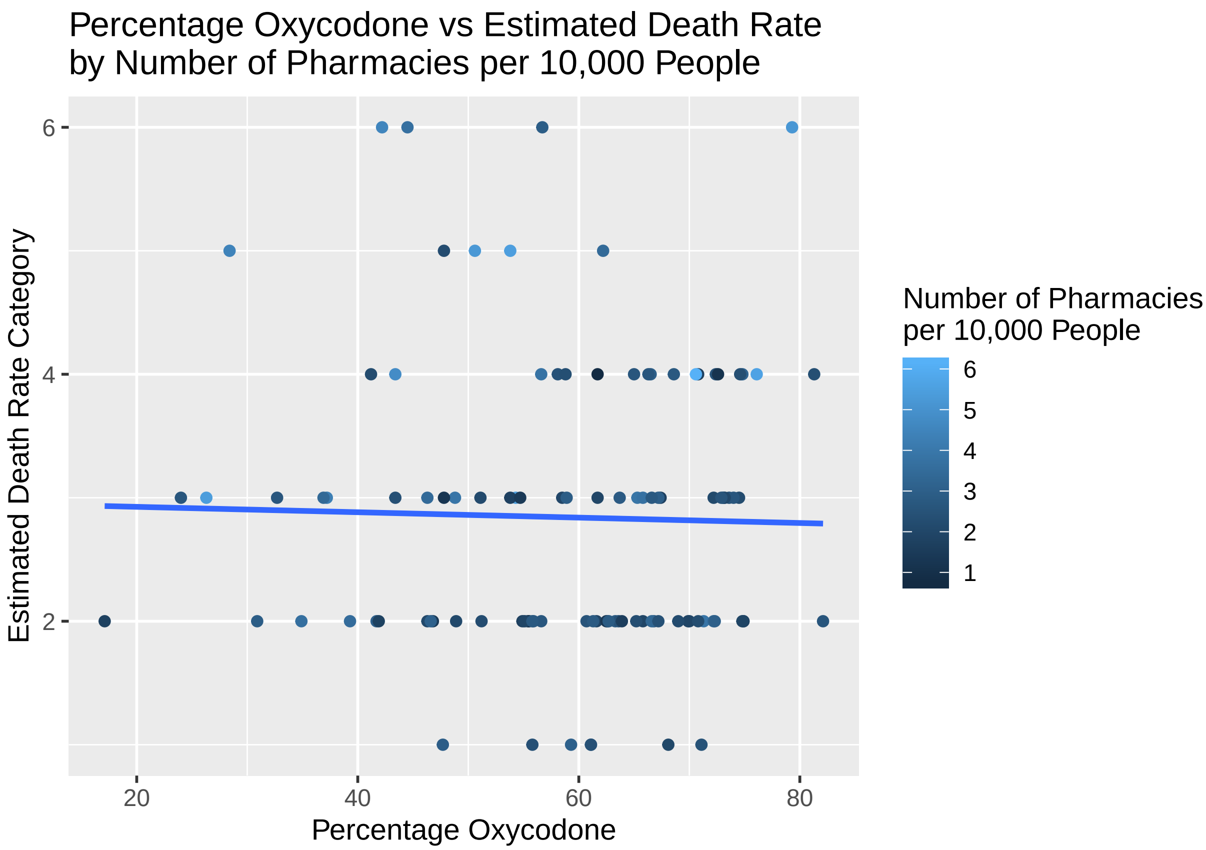
```
# number of manufactueres (ptt) vs death rate by number of pharmacies (ptt)
ggplot(train_oh_wv_2012,
       aes(x = distr_num_ptt, y = as.numeric(est_death_rate_cat), color = pharmacy_num_ptt)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Number of Distributors per 10,000 People vs Estimated Death Rate \nby Number of Pharmaci
       x = "Number of Distributors per 10,000 People",
       y = "Estimated Death Rate Category",
       color = "Number of Pharmacies \nper 10,000 People")
```



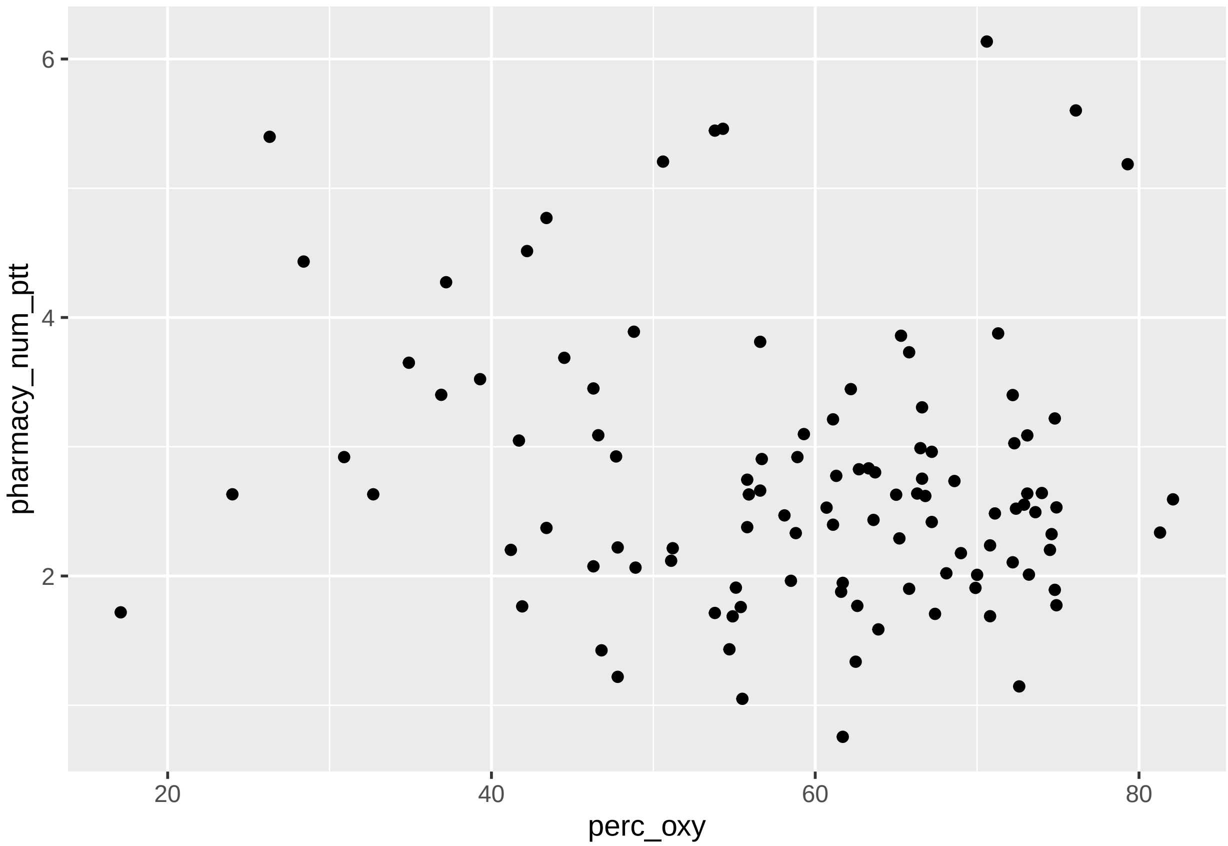

```r
ggplot(train_oh_wv_2012,
       aes(x = distr_num_ptt, y = pharmacy_num_ptt))+
  geom_point()
```



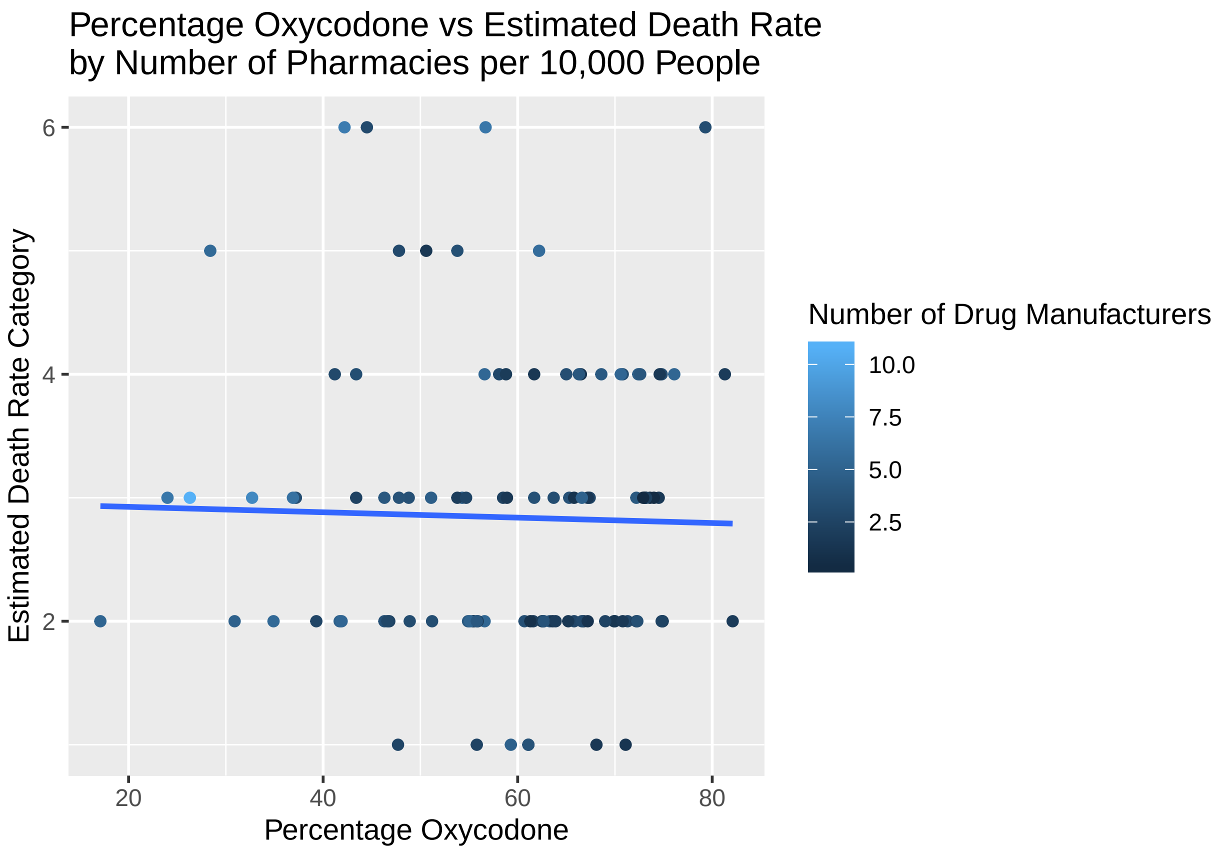

```r
# % oxycodone vs death rate by number of pharmacies (ptt)
ggplot(train_oh_wv_2012,
       aes(x = perc_oxy, y = as.numeric(est_death_rate_cat), color = pharmacy_num_ptt)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Percentage Oxycodone vs Estimated Death Rate \nby Number of Pharmacies per 10,000 People
       x = "Percentage Oxycodone",
       y = "Estimated Death Rate Category",
       color = "Number of Pharmacies \nper 10,000 People")
```



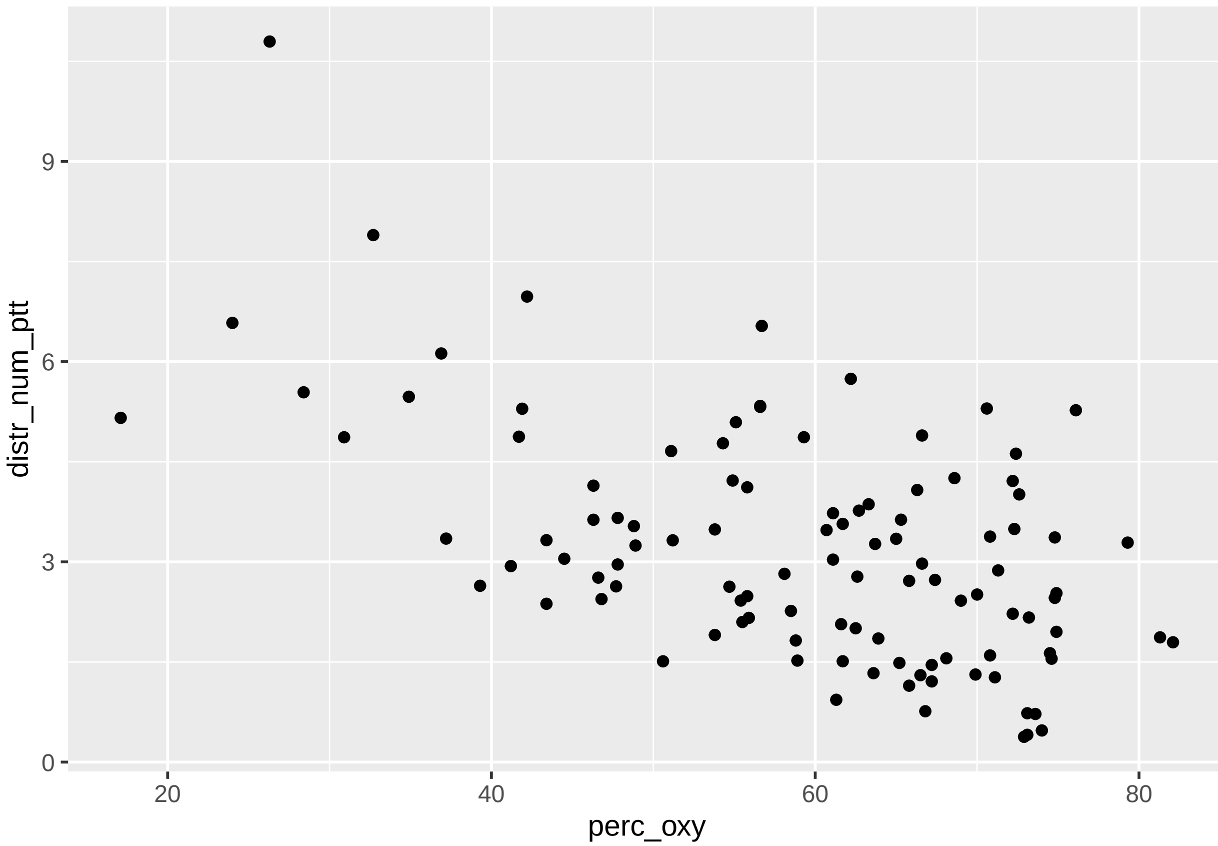

```r
ggplot(train_oh_wv_2012,
       aes(x = perc_oxy, y = pharmacy_num_ptt))+
  geom_point()
```



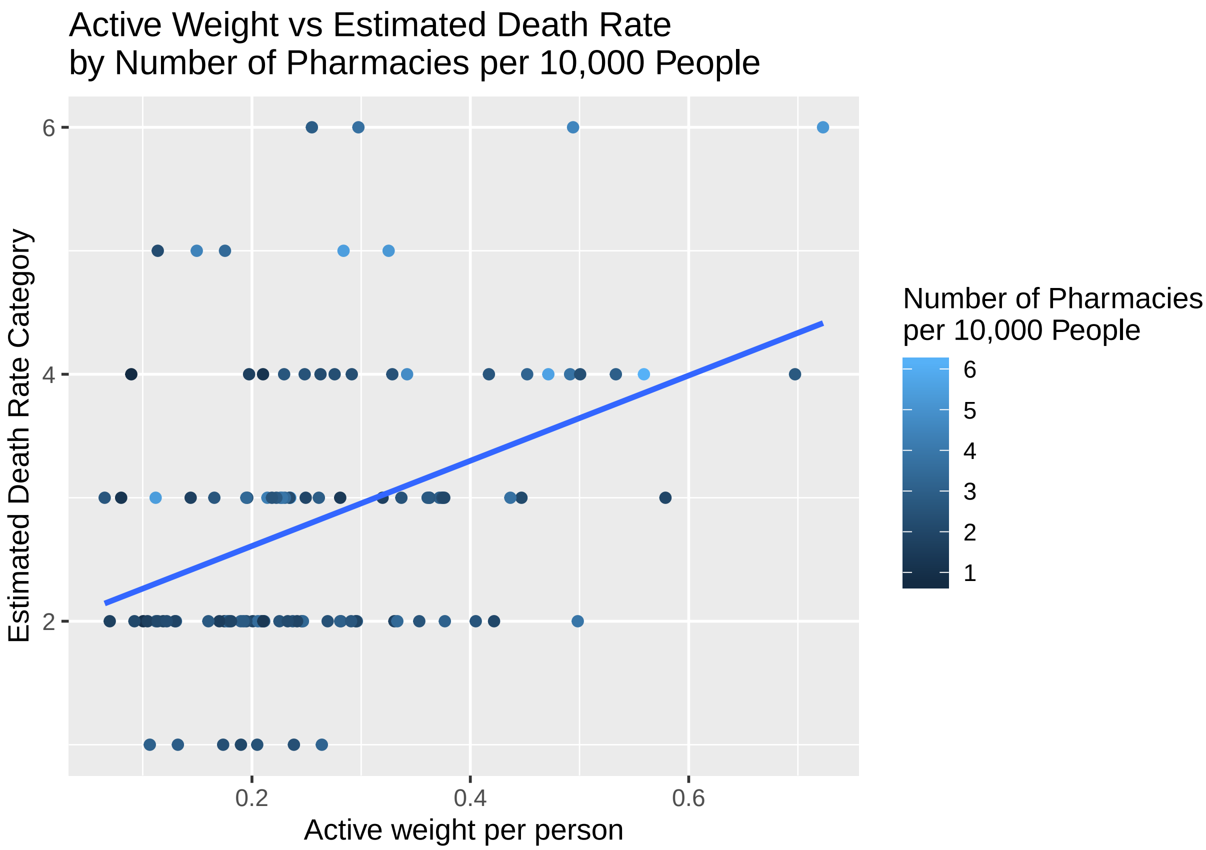

```r
# % oxycodone vs death rate by number of manufacturers
ggplot(train_oh_wv_2012,
       aes(x = perc_oxy, y = as.numeric(est_death_rate_cat), color = distr_num_ptt)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Percentage Oxycodone vs Estimated Death Rate \nby Number of Pharmacies per 10,000 People
       x = "Percentage Oxycodone",
       y = "Estimated Death Rate Category",
       color = "Number of Drug Manufacturers")
```



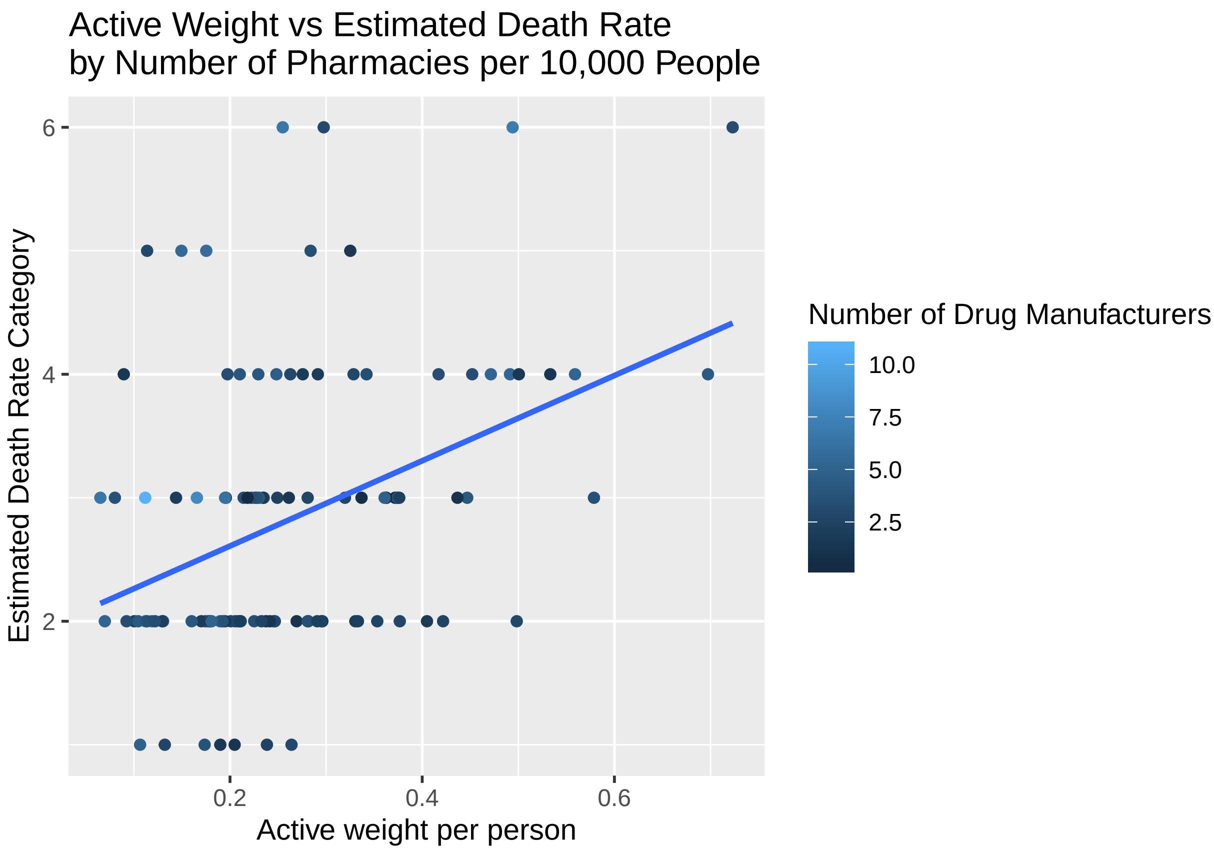

```
ggplot(train_oh_wv_2012,
       aes(x = perc_oxy, y = distr_num_ptt))+
  geom_point()
```



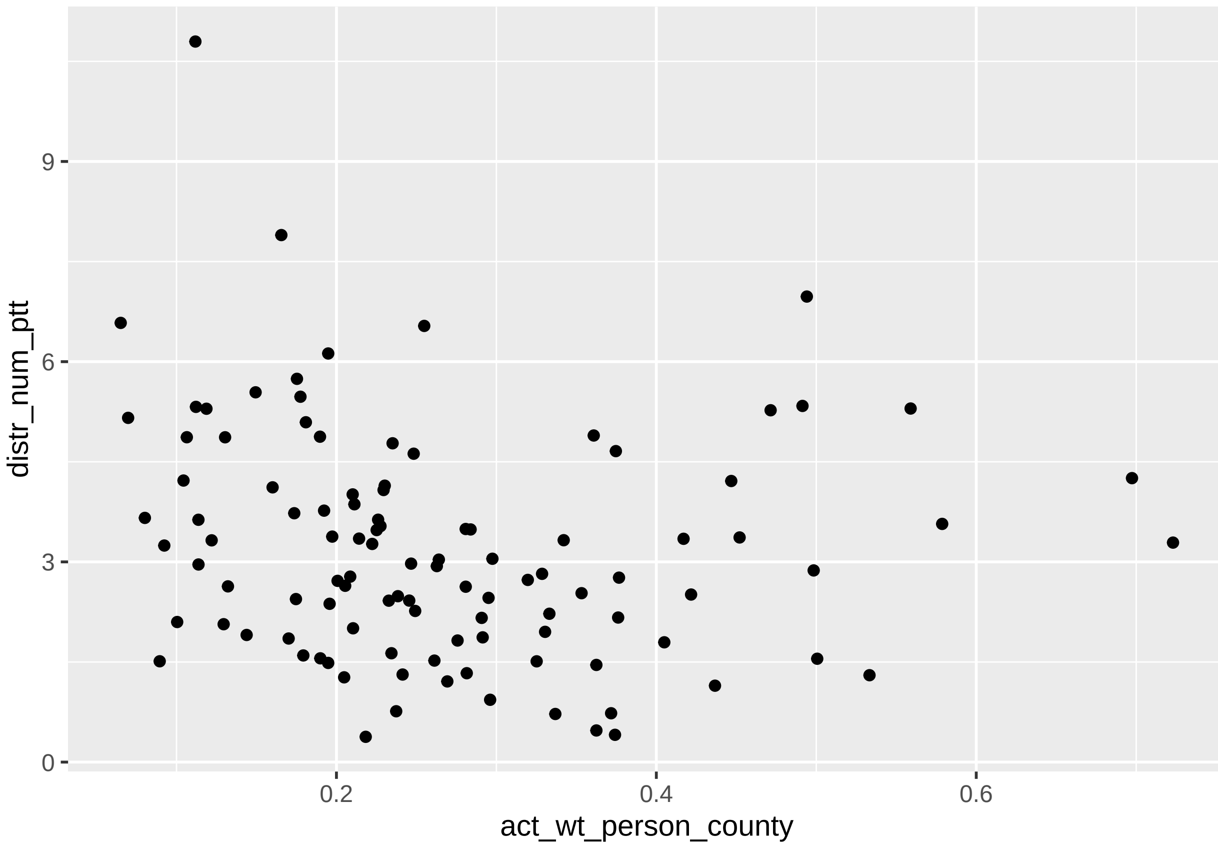

```r
# active weight per person vs number of pharmacies
ggplot(train_oh_wv_2012,
       aes(x = act_wt_person_county, y = as.numeric(est_death_rate_cat), color = pharmacy_num_ptt)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Active Weight vs Estimated Death Rate \nby Number of Pharmacies per 10,000 People ",
       x = "Active weight per person",
       y = "Estimated Death Rate Category",
       color = "Number of Pharmacies \nper 10,000 People")
```



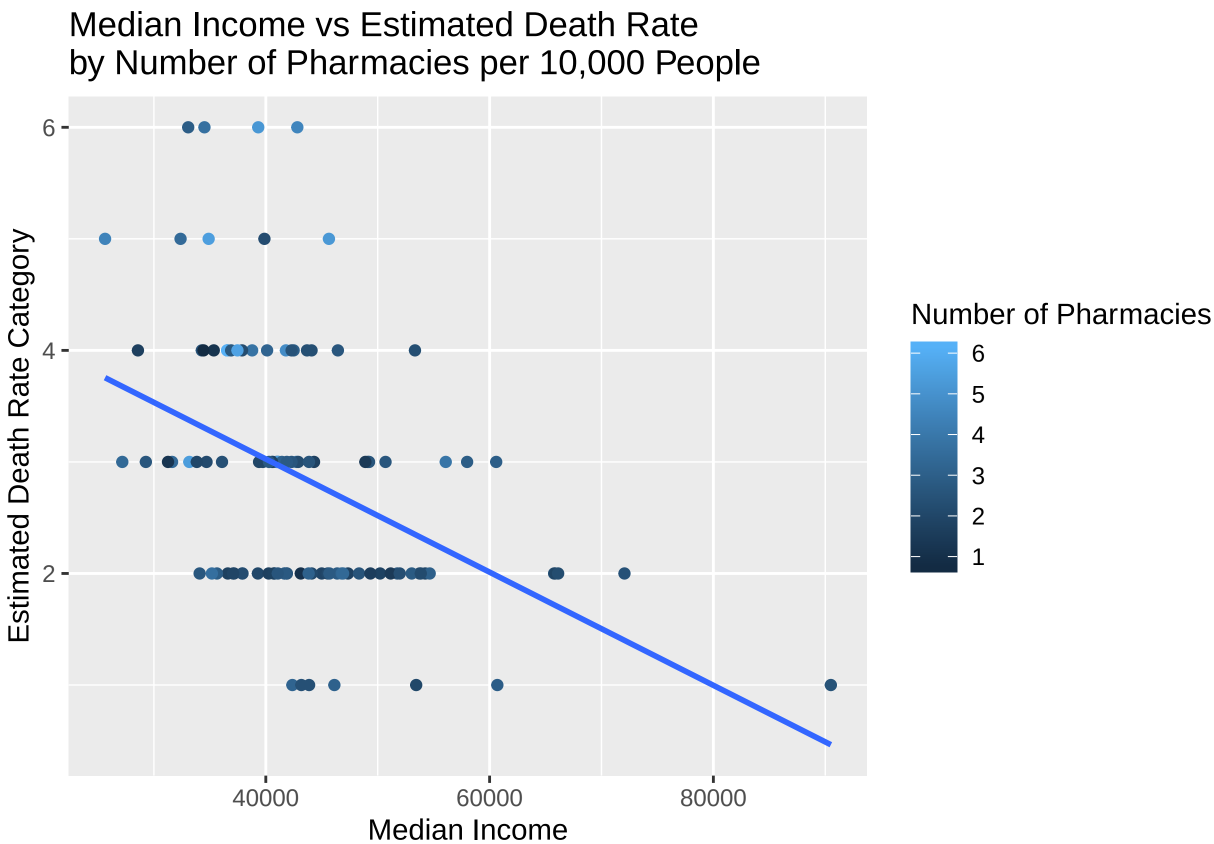

```r
# number of manufacturers vs active weight per person
ggplot(train_oh_wv_2012,
       aes(x = act_wt_person_county, y = as.numeric(est_death_rate_cat), color = distr_num_ptt)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Active Weight vs Estimated Death Rate \nby Number of Pharmacies per 10,000 People ",
       x = "Active weight per person",
       y = "Estimated Death Rate Category",
       color = "Number of Drug Manufacturers")
```



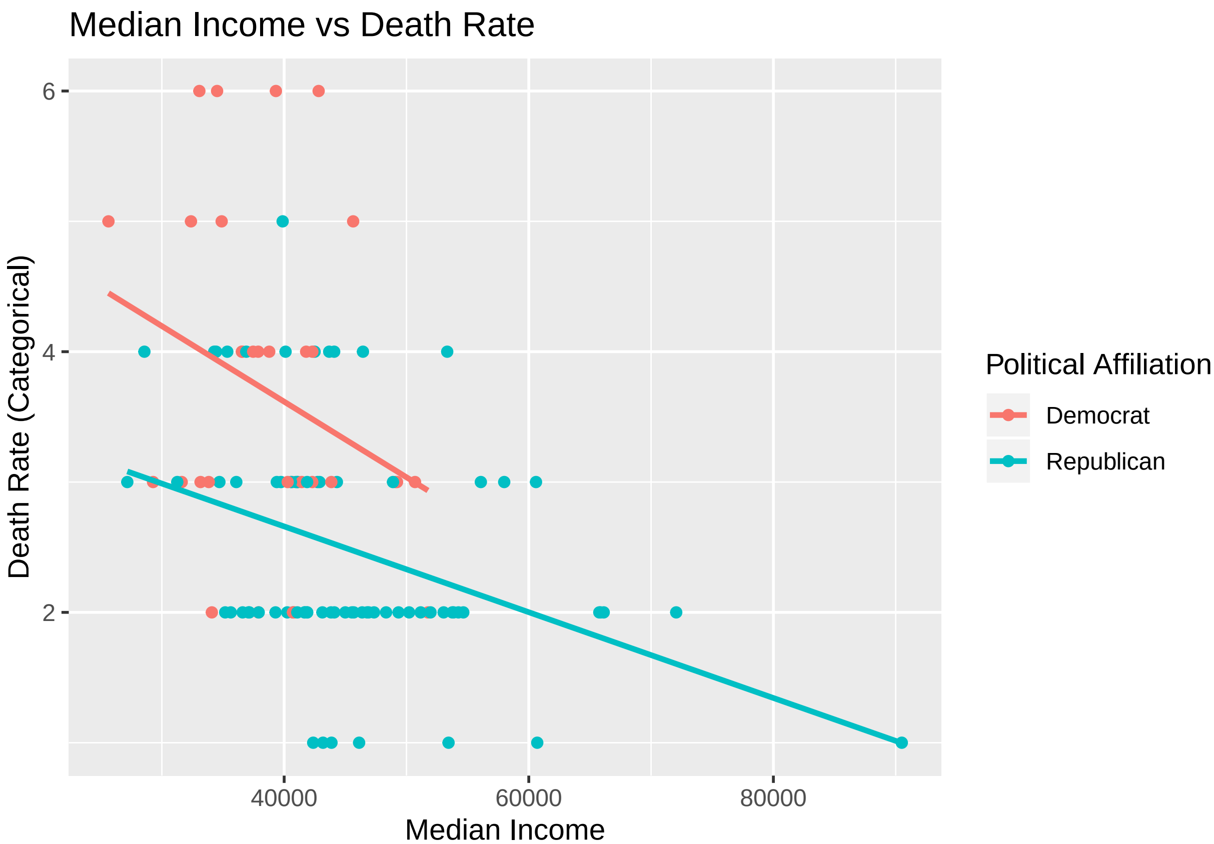

```
ggplot(train_oh_wv_2012,
       aes(x = act_wt_person_county, y = distr_num_ptt))+
  geom_point()
```



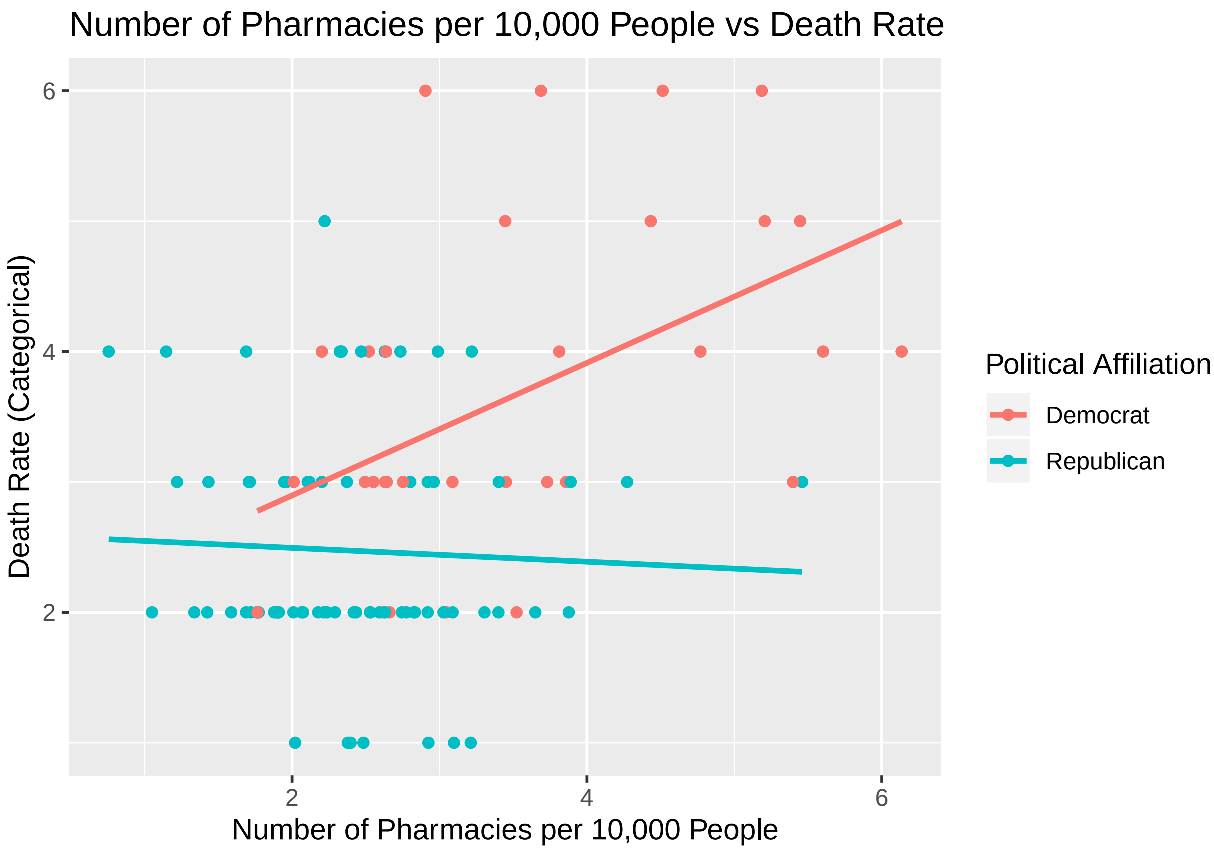

```
# income vs number of pharmacies
ggplot(train_oh_wv_2012,
       aes(x = median_income, y = as.numeric(est_death_rate_cat), color = pharmacy_num_ptt)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Median Income vs Estimated Death Rate \nby Number of Pharmacies per 10,000 People ",
       x = "Median Income",
       y = "Estimated Death Rate Category",
       color = "Number of Pharmacies")
```



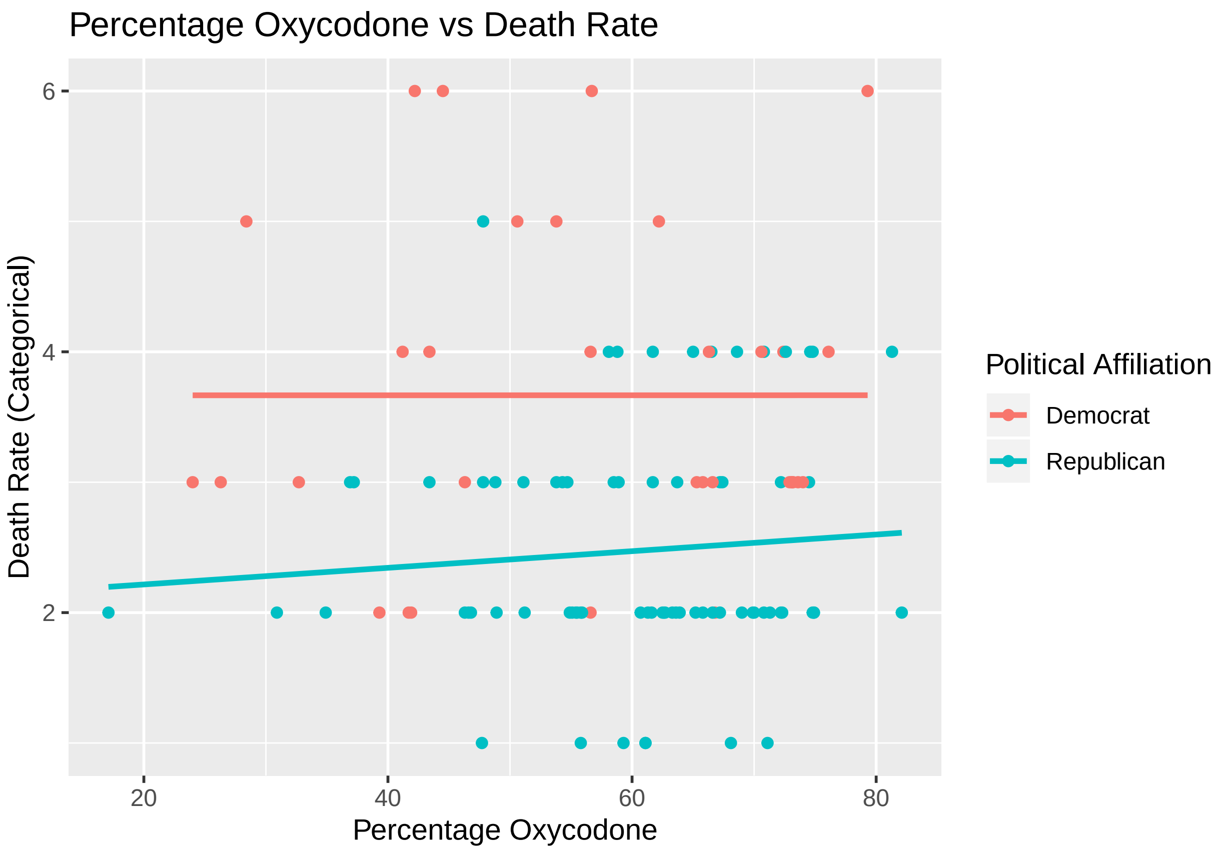

```r
# interaction b/w median income and political affiliation
ggplot(data = train_oh_wv_2012,
            aes(x = median_income ,y = as.numeric(est_death_rate_cat),color = political_aff)) +
  geom_point()+
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Median Income vs Death Rate",
       x = "Median Income", y = "Death Rate (Categorical)",
       color = "Political Affiliation")
```



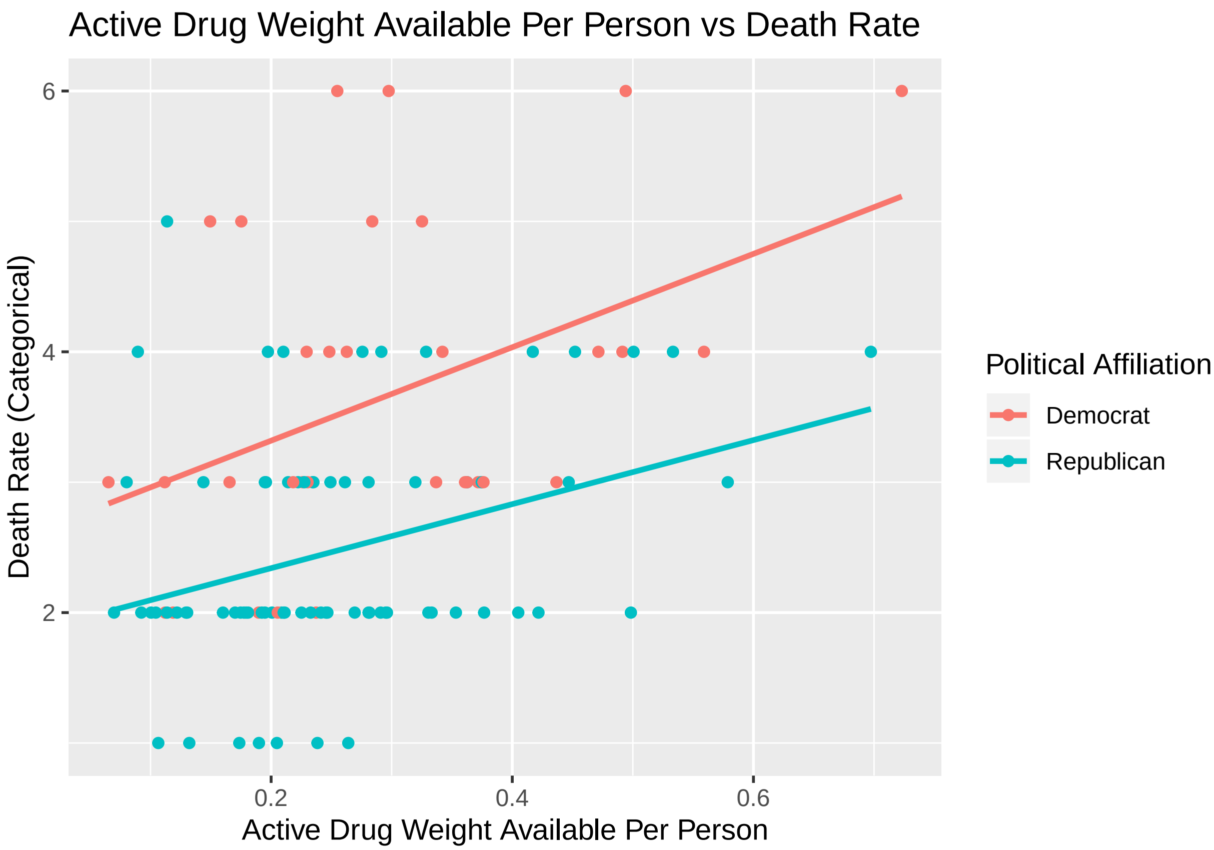

```r
# Interaction b/w number of pharmacies and political affliation
ggplot(data=train_oh_wv_2012,
       aes(x= pharmacy_num_ptt ,y = as.numeric(est_death_rate_cat), color= political_aff)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Number of Pharmacies per 10,000 People vs Death Rate",
       x = "Number of Pharmacies per 10,000 People", y = "Death Rate (Categorical)",
       color = "Political Affiliation")
```



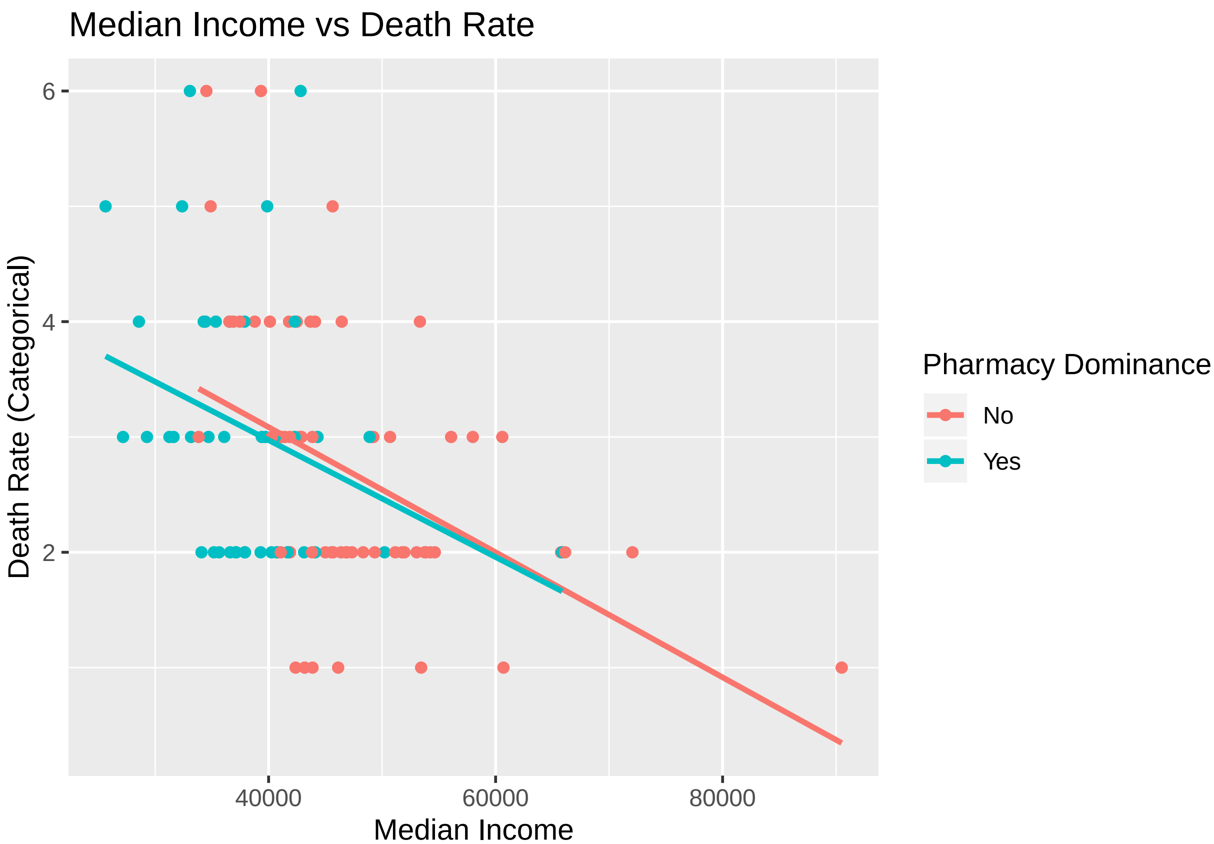

```r
# Interaction b/w % oxycodone and political affliation
ggplot(data=train_oh_wv_2012,
             aes(x = perc_oxy,y = as.numeric(est_death_rate_cat), color= political_aff)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title="Percentage Oxycodone vs Death Rate",
       x = "Percentage Oxycodone", y= "Death Rate (Categorical)",
       color ="Political Affiliation")
```



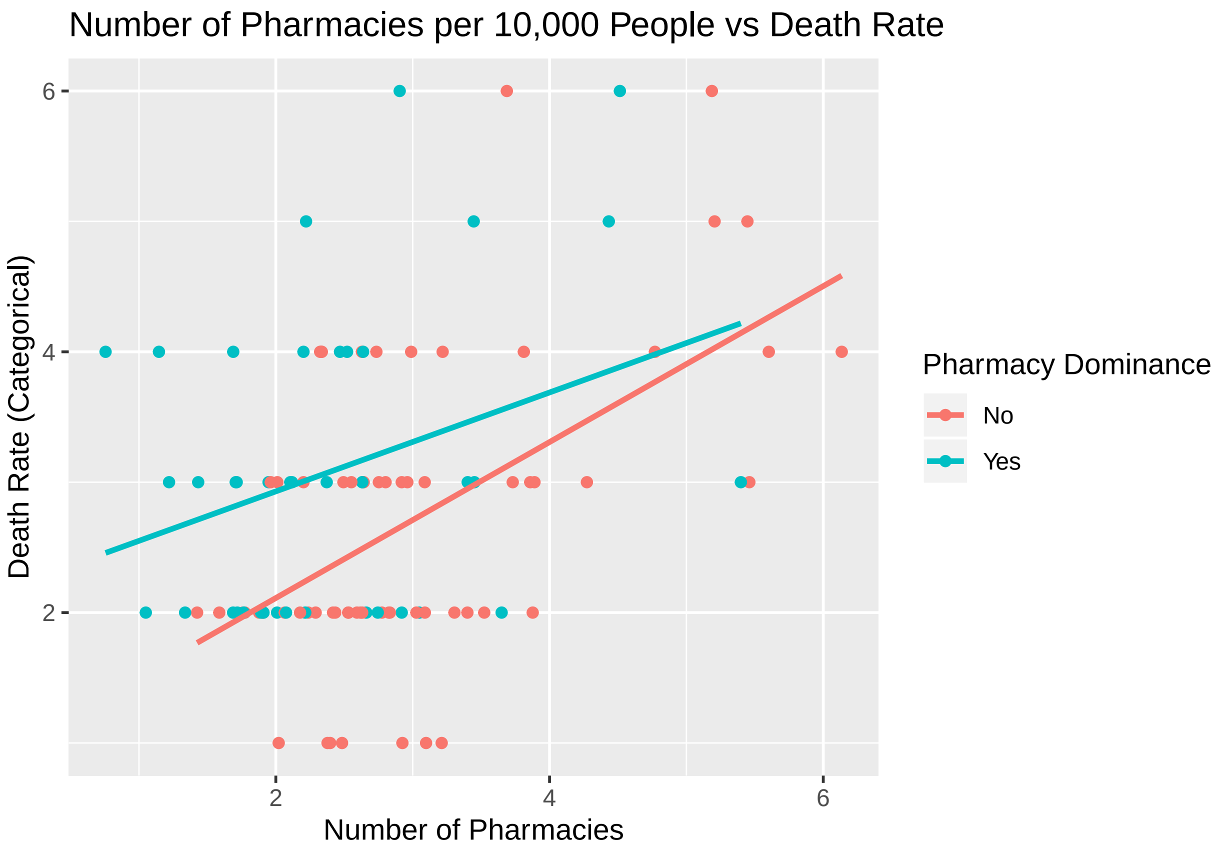

```r
# Interaction b/w active weight per person and political affliation
ggplot(data = train_oh_wv_2012,
            aes(x= act_wt_person_county ,y = as.numeric(est_death_rate_cat), color= political_aff)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title="Active Drug Weight Available Per Person vs Death Rate",
       x = "Active Drug Weight Available Per Person", y = "Death Rate (Categorical)",
       color = "Political Affiliation")
```



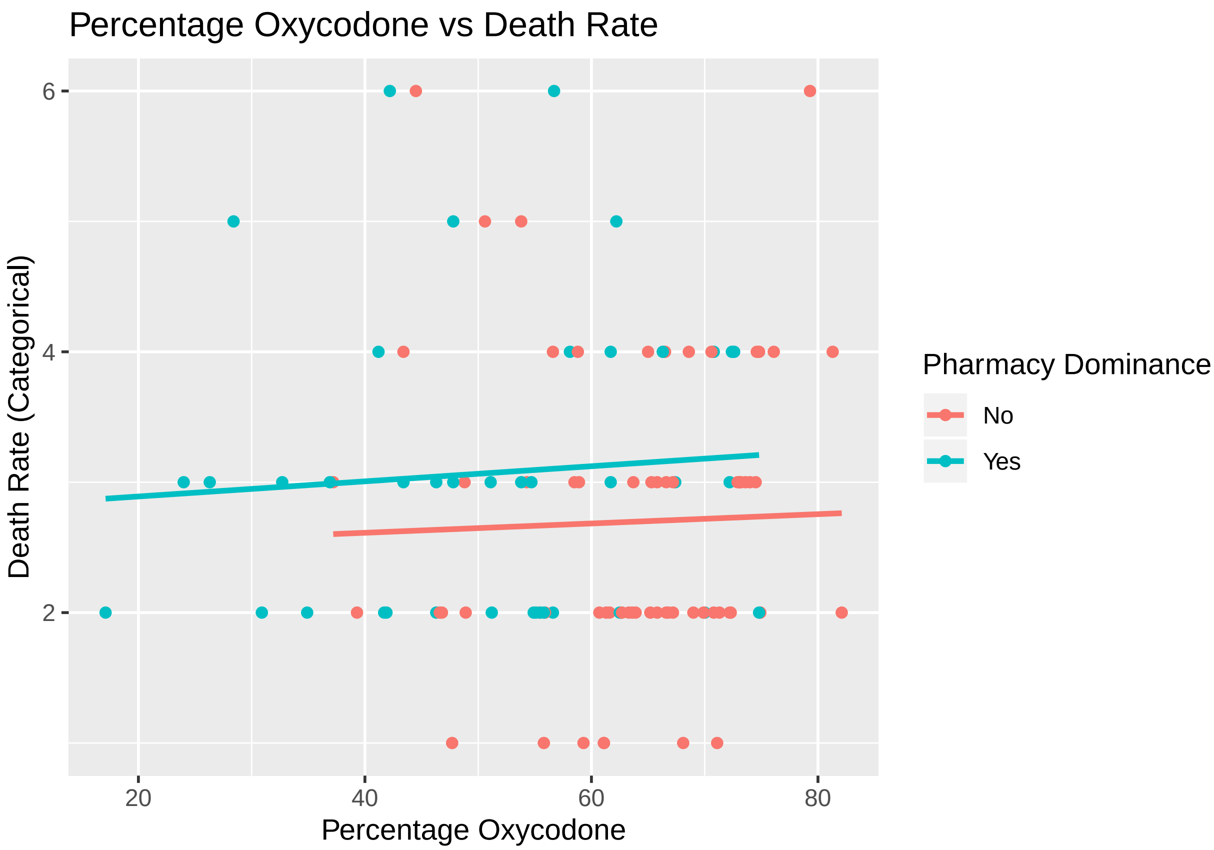

```r
# interaction b/w median income and pharmacy dominance
ggplot(data = train_oh_wv_2012,
            aes(x = median_income ,y = as.numeric(est_death_rate_cat),color = dominance)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Median Income vs Death Rate",
       x = "Median Income", y = "Death Rate (Categorical)",
       color = "Pharmacy Dominance")
```



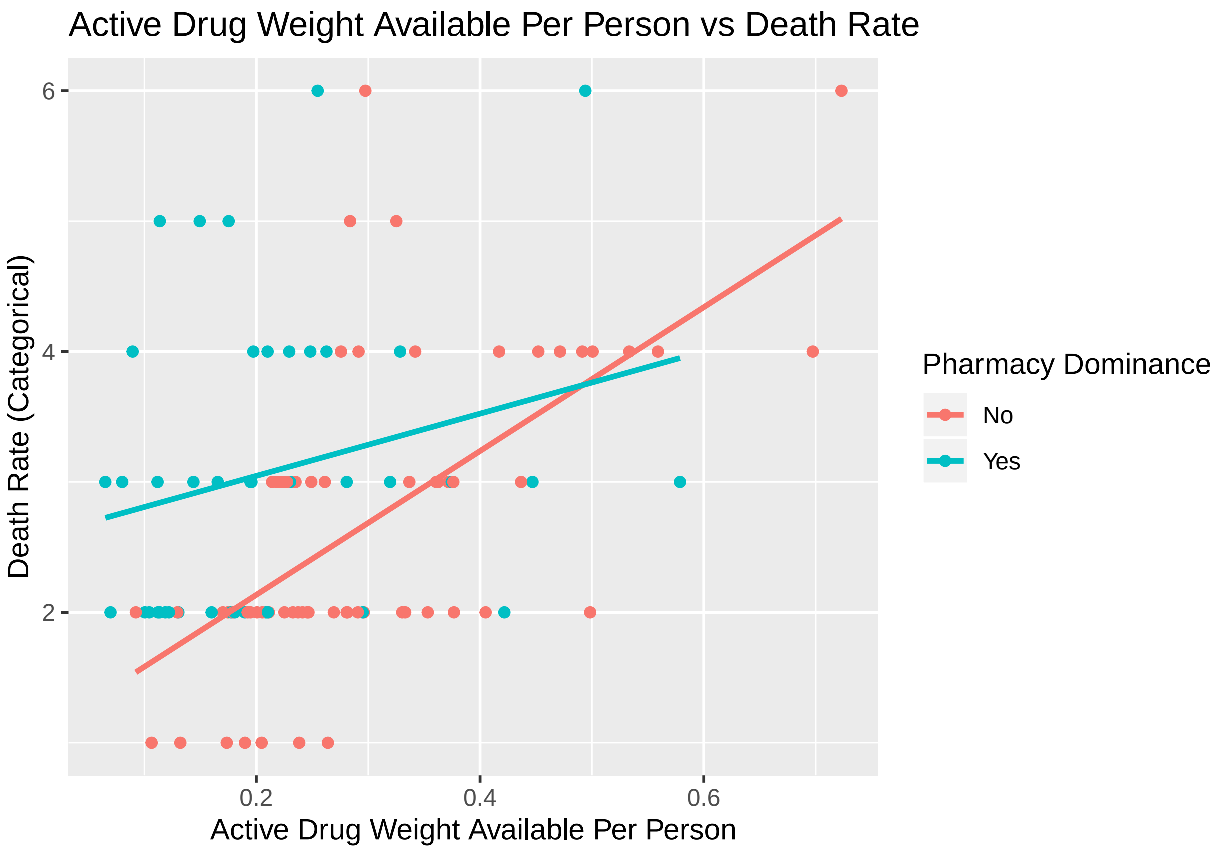

```r
# Interaction b/w number of pharmacies and pharmacy dominance
ggplot(data=train_oh_wv_2012,
       aes(x= pharmacy_num_ptt ,y=as.numeric(est_death_rate_cat) ,color= dominance)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Number of Pharmacies per 10,000 People vs Death Rate",
       x = "Number of Pharmacies", y = "Death Rate (Categorical)",
       color = "Pharmacy Dominance")
```



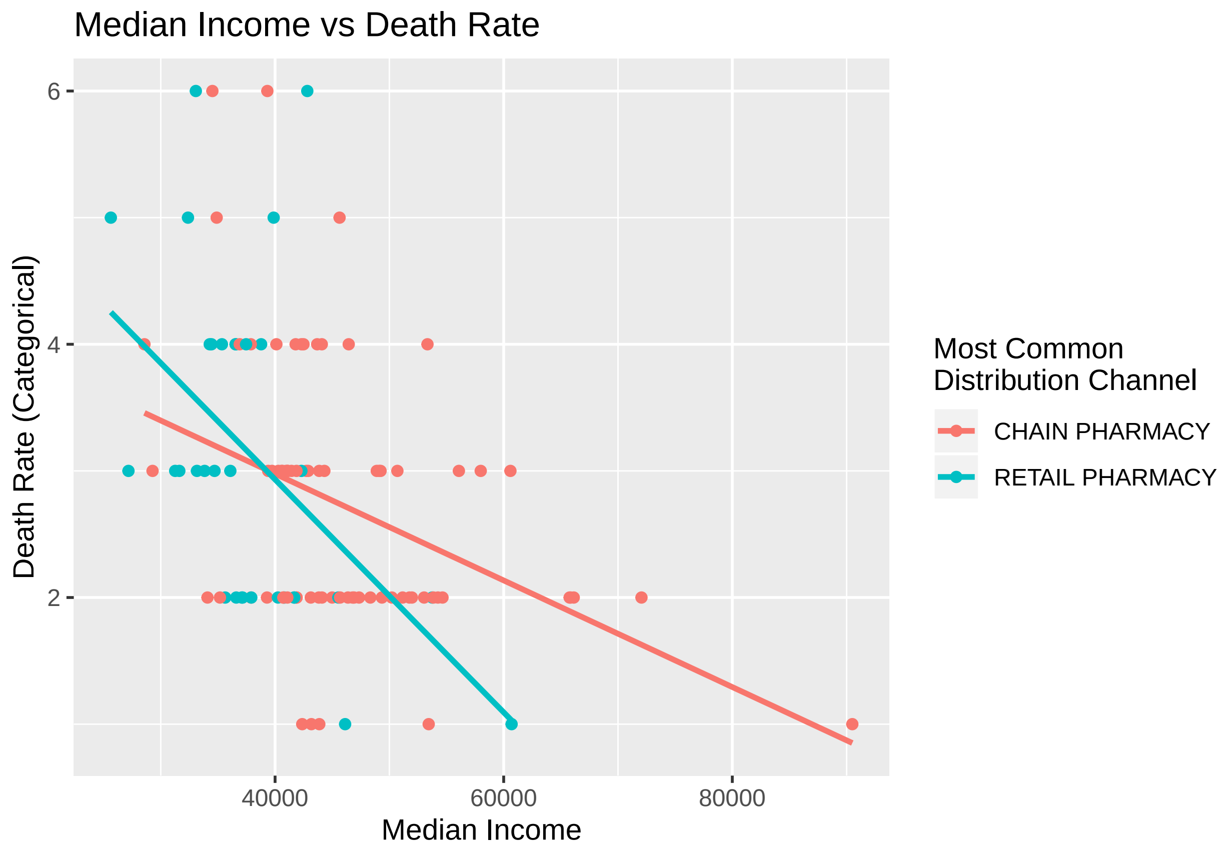

```r
# Interaction b/w % oxycodone and political affliation
ggplot(data=train_oh_wv_2012,
             aes(x = perc_oxy,y = as.numeric(est_death_rate_cat), color= dominance)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title="Percentage Oxycodone vs Death Rate",
       x = "Percentage Oxycodone", y= "Death Rate (Categorical)",
       color ="Pharmacy Dominance")
```



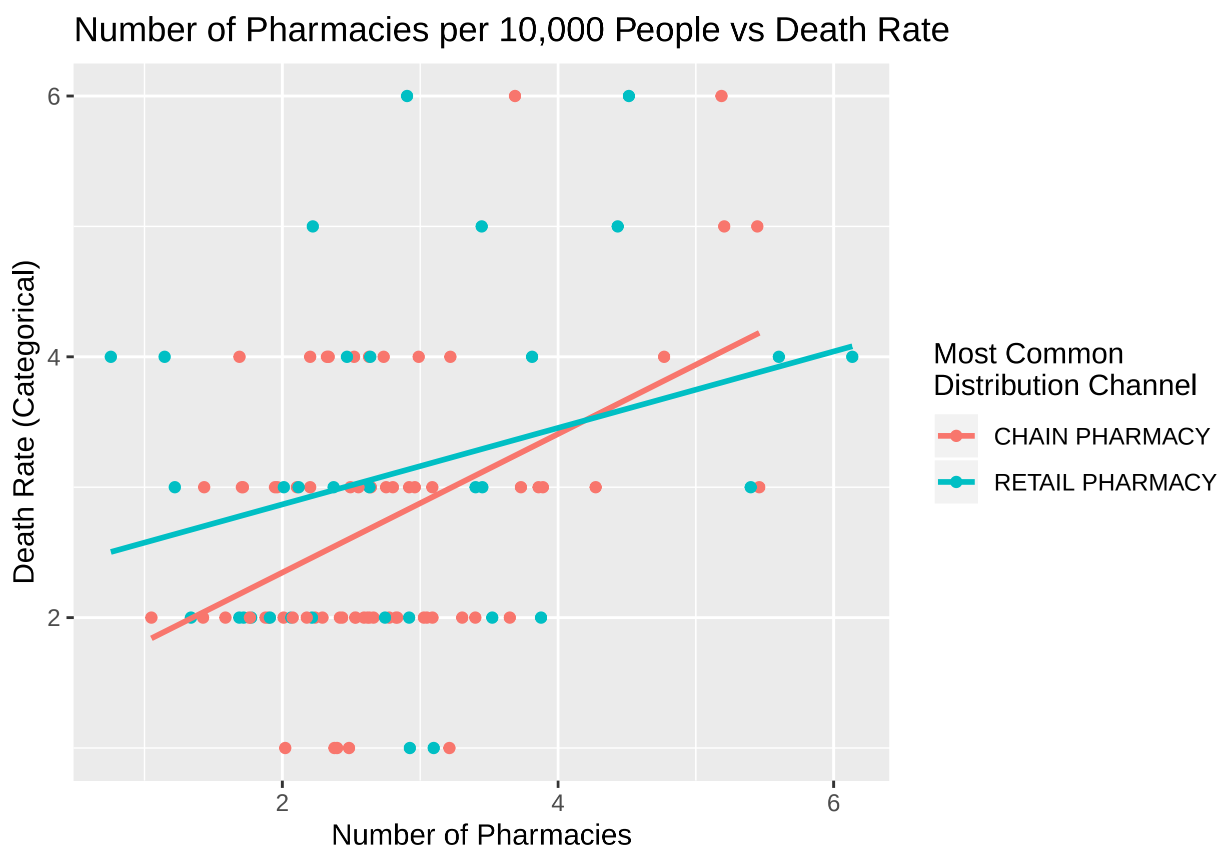

```r
# Interaction b/w active weight per person and political affliation
ggplot(data = train_oh_wv_2012,
            aes(x= act_wt_person_county ,y = as.numeric(est_death_rate_cat), color= dominance)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title="Active Drug Weight Available Per Person vs Death Rate",
       x = "Active Drug Weight Available Per Person", y = "Death Rate (Categorical)",
       color = "Pharmacy Dominance")
```



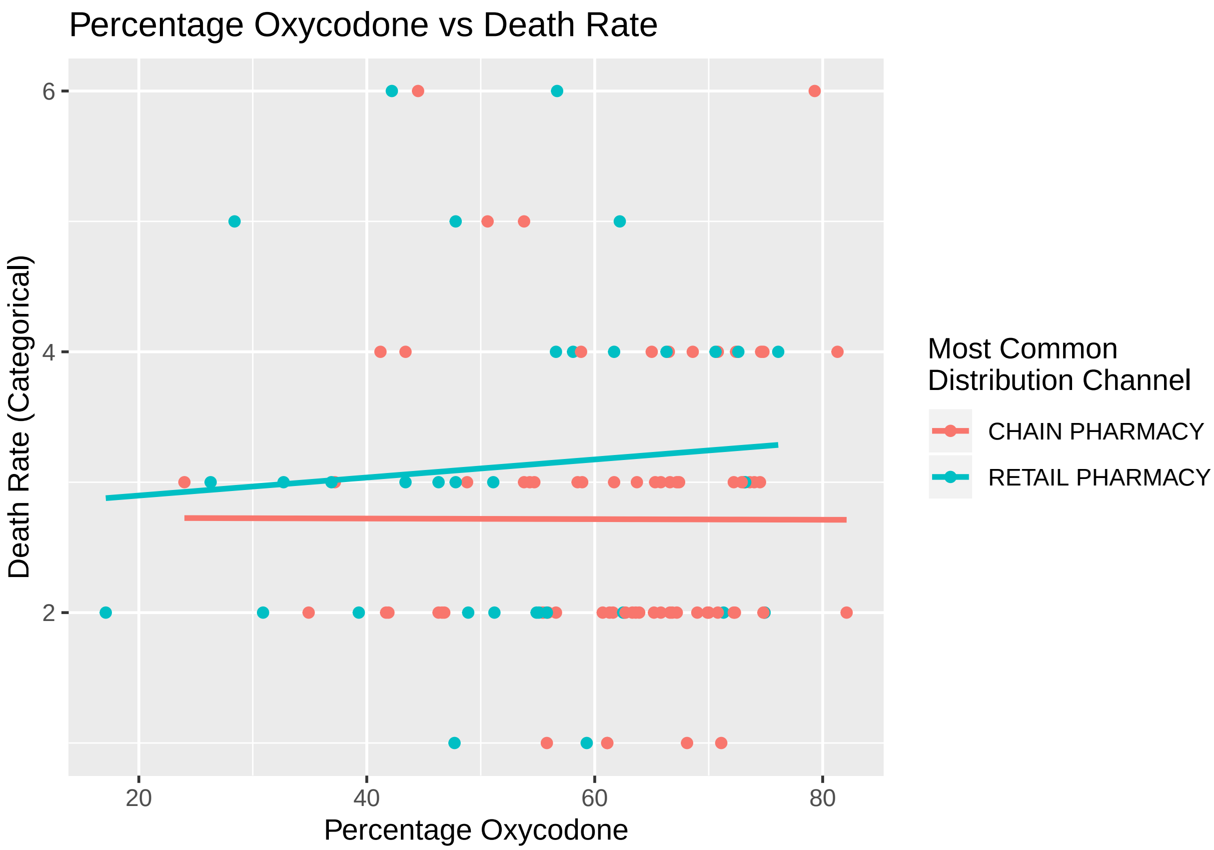

```r
# interaction b/w median income and most common distribution channel
ggplot(data = train_oh_wv_2012,
            aes(x = median_income ,y = as.numeric(est_death_rate_cat),color = most_dist_channel)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Median Income vs Death Rate",
       x = "Median Income", y = "Death Rate (Categorical)",
       color = "Most Common \nDistribution Channel")
```



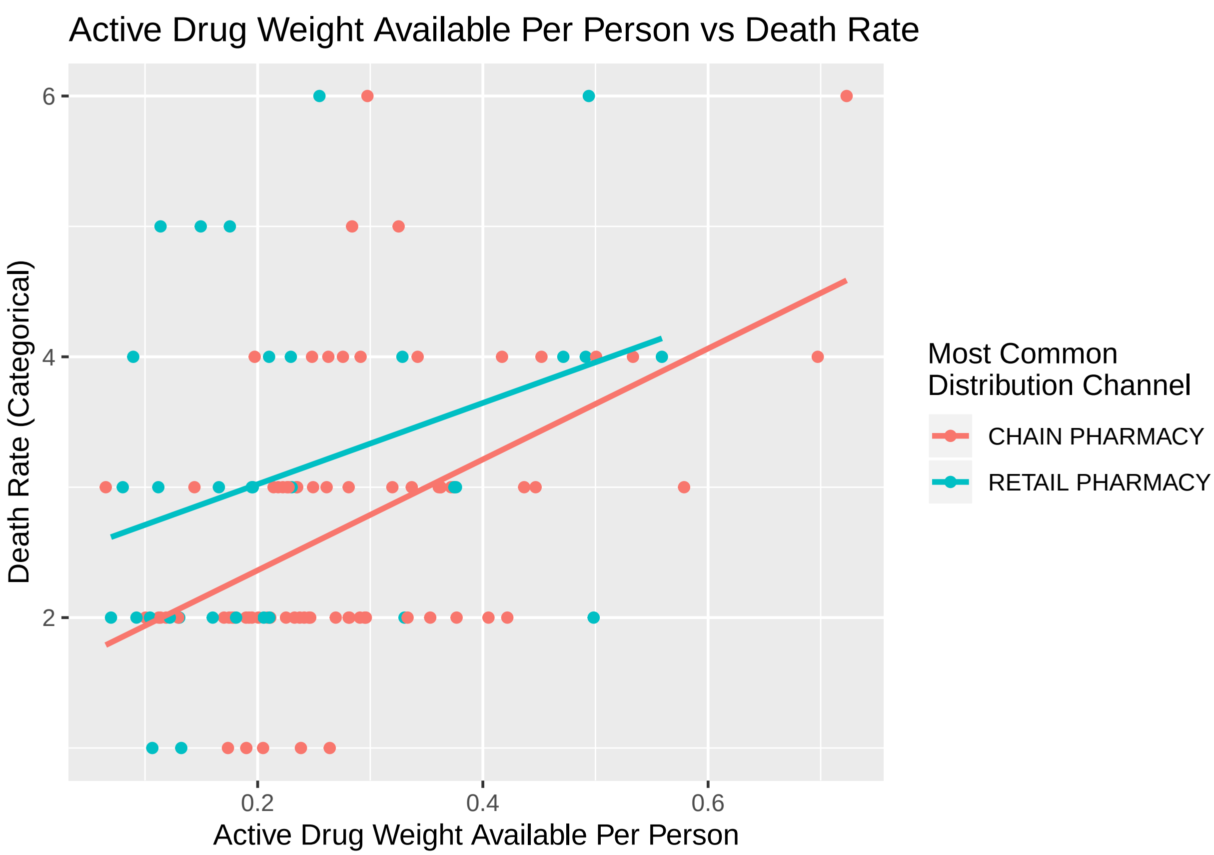

```r
# Interaction b/w number of pharmacies and most common distribtuion channel
ggplot(data=train_oh_wv_2012,
       aes(x= pharmacy_num_ptt ,y=as.numeric(est_death_rate_cat) ,color= most_dist_channel)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Number of Pharmacies per 10,000 People vs Death Rate",
       x = "Number of Pharmacies", y = "Death Rate (Categorical)",
       color = "Most Common \nDistribution Channel")
```




```
# Interaction b/w % oxycodone and most common distribtuion channel
ggplot(data=train_oh_wv_2012,
             aes(x = perc_oxy,y = as.numeric(est_death_rate_cat), color= most_dist_channel)) +
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title="Percentage Oxycodone vs Death Rate",
       x = "Percentage Oxycodone", y= "Death Rate (Categorical)",
       color ="Most Common \nDistribution Channel")
```

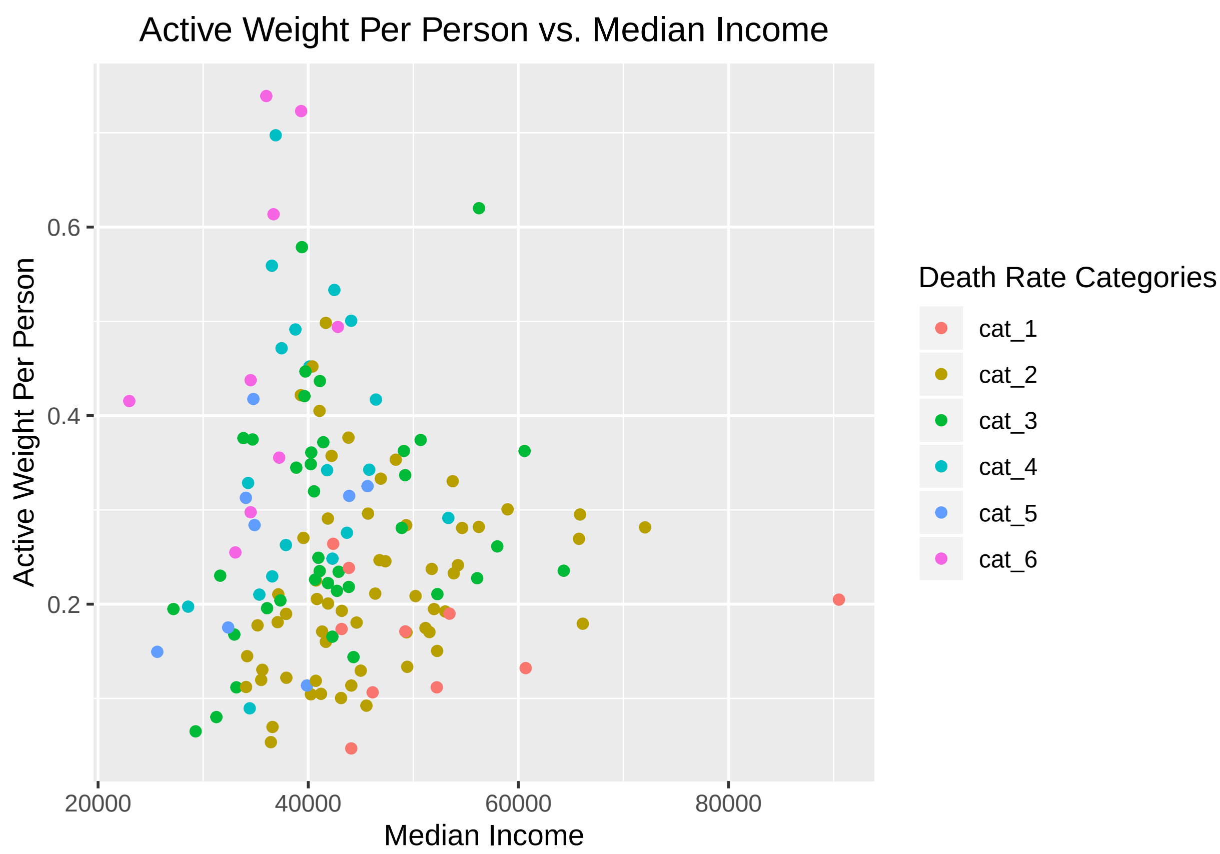



```r
# Interaction b/w active weight per person and most common distribtuion channel
ggplot(data = train_oh_wv_2012,
            aes(x= act_wt_person_county ,y = as.numeric(est_death_rate_cat), color= most_dist_channel))
  geom_point() +
  geom_smooth(se = FALSE, method = "lm") +
  labs(title="Active Drug Weight Available Per Person vs Death Rate",
       x = "Active Drug Weight Available Per Person", y = "Death Rate (Categorical)",
       color = "Most Common \nDistribution Channel")
```



Quick check for clustering

```
ggplot(oh_wv_2012,
       aes(x = median_income, y = act_wt_person_county, color = as.factor(est_death_rate_cat))) +
  # geom_mark_ellipse(aes(fill = as.factor(est_death_rate_cat))) +
  geom_point() +
  # geom_smooth(se = FALSE, method = "lm") +
  labs(title = "Active Weight Per Person vs. Median Income",
       x = "Median Income",
       y = "Active Weight Per Person",
       color = "Death Rate Categories") +
  theme(plot.title = element_text(hjust = 0.5))
```



Modeling

Logistic Regression (multinomial and cumulative logit)

Regular Multinomial Logistic Regression

```
# without interactions
fit0 <- nnet::multinom(est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance + log(medi
```

```
## # weights:  60 (45 variable)
## initial  value 191.718263
## iter  10 value 140.792988
## iter  20 value 119.719917
## iter  30 value 106.658718
## iter  40 value 101.467484
## iter  50 value 99.557245
## iter  60 value 99.193133
## iter  70 value 98.575161
## iter  80 value 97.815051
## iter  90 value 96.814965
## iter 100 value 96.633144
## final  value 96.633144
## stopped after 100 iterations
```

```r
summary(fit0)
```

```
## Call:
## nnet::multinom(formula = est_death_rate_cat ~ pharmacy_num_ptt +
##     most_dist_channel + dominance + log(median_income) + political_aff +
##     act_wt_person_county + perc_oxy + distr_num_ptt, data = train_oh_wv_2012)
##
## Coefficients:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2  -12.934470       -0.8408980                       -0.8797333
## cat_3   22.677008       -0.3687500                       -1.2249931
## cat_4   84.838577       -0.4318392                       -0.9072072
## cat_5   -7.154282       32.1402641                        5.0980038
## cat_6  100.321832       -0.8805394                       -2.4296485
##       dominanceYes log(median_income) political_affRepublican
## cat_2     43.96234           1.612957               -3.347039
## cat_3     44.70750          -1.807718               -4.671279
## cat_4     44.19591          -8.205316               -4.399158
## cat_5    102.47233         -10.067789               -4.619204
## cat_6     46.28237          -9.750815              -44.325028
##       act_wt_person_county    perc_oxy distr_num_ptt
## cat_2             16.95034 -0.03658190     0.3740006
## cat_3             22.63212 -0.04537674     0.1262562
## cat_4             25.61877  0.02035476     0.1929911
## cat_5           -187.21566  1.07498631   -15.6417048
## cat_6             41.68111 -0.07359185     0.5082274
##
## Std. Errors:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2  11.7985435        0.8665793                         1.590525
## cat_3  11.2455801        0.9036513                         1.723796
## cat_4   2.3420484        0.9919433                         1.817140
## cat_5   0.4580507        7.2860578                        16.608206
## cat_6   0.9348917        1.3020295                         2.547685
##       dominanceYes log(median_income) political_affRepublican
## cat_2     1.074216          1.2857957            5.767891e+00
## cat_3     1.002989          1.2635570            5.756633e+00
## cat_4     1.136481          0.7851238            5.787253e+00
## cat_5     2.803456          4.0874632            7.504504e+00
## cat_6     1.854378          0.8727372            5.300930e-13
##       act_wt_person_county   perc_oxy distr_num_ptt
## cat_2            10.163223 0.06695960     0.6359846
## cat_3            10.434254 0.06996255     0.6581683
## cat_4            10.643502 0.07591752     0.6978330
## cat_5             3.237991 0.39918895     4.4661326
## cat_6            12.827994 0.09859824     0.8677107
##
## Residual Deviance: 193.2663
## AIC: 283.2663
```

```r
# with interactions
fit0.interact <- nnet::multinom(est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance +
```

```
## # weights:  108 (85 variable)
```

```
## initial  value 191.718263
## iter  10 value 131.073412
## iter  20 value 118.096117
## iter  30 value 110.830122
## iter  40 value 98.770676
## iter  50 value 90.942099
## iter  60 value 83.536417
## iter  70 value 79.852409
## iter  80 value 76.558644
## iter  90 value 75.388227
## iter 100 value 73.332613
## final  value 73.332613
## stopped after 100 iterations
```

```r
summary(fit0.interact)
```

```
## Call:
## nnet::multinom(formula = est_death_rate_cat ~ pharmacy_num_ptt +
##     most_dist_channel + dominance + log(median_income) + political_aff +
##     act_wt_person_county + perc_oxy + distr_num_ptt + log(median_income) *
##     political_aff + act_wt_person_county * distr_num_ptt + act_wt_person_county *
##     pharmacy_num_ptt + pharmacy_num_ptt * political_aff + log(median_income) *
##     most_dist_channel + log(median_income) * pharmacy_num_ptt +
##     perc_oxy * dominance + act_wt_person_county * perc_oxy, data = train_oh_wv_2012)
##
## Coefficients:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2  -70.039122         19.30373                       -89.134340
## cat_3   57.272614        -55.95872                        78.408428
## cat_4   13.254524        -19.41151                        18.793324
## cat_5    2.310263         33.15582                         4.226015
## cat_6   36.248195        112.81148                        11.483641
##       dominanceYes log(median_income) political_affRepublican
## cat_2     7.902782          19.782504              -23.925745
## cat_3    11.601608           6.686706               61.141209
## cat_4    15.459828           8.850545                4.377992
## cat_5    70.423648         -12.318535               -7.510619
## cat_6   -85.515095          17.166640               -4.394870
##       act_wt_person_county   perc_oxy distr_num_ptt
## cat_2           -46.379122 -0.1179295      2.861000
## cat_3            -6.067798 -0.1091219      2.654085
## cat_4            43.555643  0.3700651      3.690482
## cat_5           -22.441417  2.3399420    -25.874907
## cat_6             8.747659 -2.0718575      7.930411
##       log(median_income):political_affRepublican
## cat_2                                  -9.722467
## cat_3                                 -17.581315
## cat_4                                 -12.334707
## cat_5                                  -8.003669
## cat_6                                 -15.001181
##       act_wt_person_county:distr_num_ptt
## cat_2                          -12.40656
## cat_3                          -11.63255
## cat_4                          -13.87518
## cat_5                          -23.17532
```

```
## cat_6                          -23.99092
##       pharmacy_num_ptt:act_wt_person_county
## cat_2                              27.45755
## cat_3                              15.18193
## cat_4                              34.42985
## cat_5                             -54.60156
## cat_6                              59.77132
##       pharmacy_num_ptt:political_affRepublican
## cat_2                                18.065213
## cat_3                                16.777019
## cat_4                                17.387340
## cat_5                                 2.975752
## cat_6                                13.994979
##       most_dist_channelRETAIL PHARMACY:log(median_income)
## cat_2                                               8.118516
## cat_3                                              -7.882542
## cat_4                                              -2.252188
## cat_5                                               1.155831
## cat_6                                              -1.501222
##       pharmacy_num_ptt:log(median_income) dominanceYes:perc_oxy
## cat_2                          -4.1193382             0.8982226
## cat_3                           3.3691941             0.8442034
## cat_4                          -0.6501612             0.7953694
## cat_5                           3.1198174             1.0870959
## cat_6                         -14.3072082             2.8972880
##       act_wt_person_county:perc_oxy
## cat_2                     0.4628329
## cat_3                     0.3988681
## cat_4                    -0.9349026
## cat_5                    -5.6523966
## cat_6                     0.6433017
##
## Std. Errors:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2   2.6650838        6.5507795                        0.3214459
## cat_3   3.0114917        7.9791174                        0.3058579
## cat_4   4.8929145       11.7736173                        0.5091622
## cat_5   0.6960112        0.3705614                        0.6826628
## cat_6   0.2809398        0.3182387                        0.1076812
##       dominanceYes log(median_income) political_affRepublican
## cat_2    2.9929066           1.833568            2.801574e+00
## cat_3    2.8652957           1.849142            2.923999e+00
## cat_4    4.0429297           1.868611            4.186294e+00
## cat_5    0.6825959           7.489523            5.637889e-01
## cat_6    0.7120080           2.927212            7.424363e-06
##       act_wt_person_county  perc_oxy distr_num_ptt
## cat_2           13.0786531 0.1122101     0.9950559
## cat_3           13.7169841 0.1266594     0.9734352
## cat_4            1.7099773 0.1390232     1.1445612
## cat_5            0.3470802 2.6505067     5.8107922
## cat_6            0.5262090 0.4412734     3.1089603
##       log(median_income):political_affRepublican
## cat_2                                1.904094e+00
## cat_3                                1.905924e+00
```

```
## cat_4                              1.951306e+00
## cat_5                              5.315860e+00
## cat_6                              7.856089e-05
##       act_wt_person_county:distr_num_ptt
## cat_2                          3.010011
## cat_3                          2.254038
## cat_4                          2.612740
## cat_5                          1.218282
## cat_6                          4.762232
##       pharmacy_num_ptt:act_wt_person_county
## cat_2                            3.8568923
## cat_3                            4.2597064
## cat_4                            4.3447002
## cat_5                            1.0087462
## cat_6                            0.9246094
##       pharmacy_num_ptt:political_affRepublican
## cat_2                            5.824328e+00
## cat_3                            5.853657e+00
## cat_4                            5.925844e+00
## cat_5                            4.425654e+00
## cat_6                            1.445569e-05
##       most_dist_channelRETAIL PHARMACY:log(median_income)
## cat_2                                          0.1755340
## cat_3                                          0.1952513
## cat_4                                          0.2154127
## cat_5                                          7.3046555
## cat_6                                          0.7158730
##       pharmacy_num_ptt:log(median_income) dominanceYes:perc_oxy
## cat_2                           1.0349992             0.4540956
## cat_3                           1.1998689             0.4539692
## cat_4                           1.5674393             0.4567686
## cat_5                           4.5805695             1.8063398
## cat_6                           0.5875612             0.4623058
##       act_wt_person_county:perc_oxy
## cat_2                     0.2798651
## cat_3                     0.3298717
## cat_4                     0.3010169
## cat_5                    13.6831843
## cat_6                     0.5922605
##
## Residual Deviance: 146.6652
## AIC: 316.6652
```

```
# fit0 and cumu.logistic are the SAME
############################################################################
cumu.logistic = nnet::multinom(est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance + l
```

```
## # weights:  60 (45 variable)
## initial  value 191.718263
## iter  10 value 140.792988
## iter  20 value 119.719917
## iter  30 value 106.658718
## iter  40 value 101.467484
## iter  50 value 99.557245
## iter  60 value 99.193133
```

```
## iter  70 value 98.575161
## iter  80 value 97.815051
## iter  90 value 96.814965
## iter 100 value 96.633144
## final  value 96.633144
## stopped after 100 iterations
```

```r
summary(cumu.logistic)
```

```
## Call:
## nnet::multinom(formula = est_death_rate_cat ~ pharmacy_num_ptt +
##     most_dist_channel + dominance + log(median_income) + political_aff +
##     act_wt_person_county + perc_oxy + distr_num_ptt, data = train_oh_wv_2012)
##
## Coefficients:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2  -12.934470       -0.8408980                       -0.8797333
## cat_3   22.677008       -0.3687500                       -1.2249931
## cat_4   84.838577       -0.4318392                       -0.9072072
## cat_5   -7.154282       32.1402641                        5.0980038
## cat_6  100.321832       -0.8805394                       -2.4296485
##       dominanceYes log(median_income) political_affRepublican
## cat_2     43.96234           1.612957               -3.347039
## cat_3     44.70750          -1.807718               -4.671279
## cat_4     44.19591          -8.205316               -4.399158
## cat_5    102.47233         -10.067789               -4.619204
## cat_6     46.28237          -9.750815              -44.325028
##       act_wt_person_county    perc_oxy distr_num_ptt
## cat_2             16.95034 -0.03658190     0.3740006
## cat_3             22.63212 -0.04537674     0.1262562
## cat_4             25.61877  0.02035476     0.1929911
## cat_5           -187.21566  1.07498631   -15.6417048
## cat_6             41.68111 -0.07359185     0.5082274
##
## Std. Errors:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2  11.7985435        0.8665793                         1.590525
## cat_3  11.2455801        0.9036513                         1.723796
## cat_4   2.3420484        0.9919433                         1.817140
## cat_5   0.4580507        7.2860578                        16.608206
## cat_6   0.9348917        1.3020295                         2.547685
##       dominanceYes log(median_income) political_affRepublican
## cat_2     1.074216          1.2857957            5.767891e+00
## cat_3     1.002989          1.2635570            5.756633e+00
## cat_4     1.136481          0.7851238            5.787253e+00
## cat_5     2.803456          4.0874632            7.504504e+00
## cat_6     1.854378          0.8727372            5.300930e-13
##       act_wt_person_county   perc_oxy distr_num_ptt
## cat_2            10.163223 0.06695960     0.6359846
## cat_3            10.434254 0.06996255     0.6581683
## cat_4            10.643502 0.07591752     0.6978330
## cat_5             3.237991 0.39918895     4.4661326
## cat_6            12.827994 0.09859824     0.8677107
##
## Residual Deviance: 193.2663
```

```
## AIC: 283.2663
```

```r
mostImportantVariables <- varImp(cumu.logistic)
mostImportantVariables$Variables <- row.names(mostImportantVariables)
mostImportantVariables <- mostImportantVariables[order(-mostImportantVariables$Overall),]
print(mostImportantVariables)
```

```
##                                        Overall
## act_wt_person_county             294.098005
## dominanceYes                     281.620449
## political_affRepublican           61.361708
## pharmacy_num_ptt                  34.662291
## log(median_income)                31.444596
## distr_num_ptt                     16.843180
## most_dist_channelRETAIL PHARMACY  10.539586
## perc_oxy                           1.250892
##                                                         Variables
## act_wt_person_county                         act_wt_person_county
## dominanceYes                                         dominanceYes
## political_affRepublican                   political_affRepublican
## pharmacy_num_ptt                                 pharmacy_num_ptt
## log(median_income)                             log(median_income)
## distr_num_ptt                                       distr_num_ptt
## most_dist_channelRETAIL PHARMACY most_dist_channelRETAIL PHARMACY
## perc_oxy                                                 perc_oxy
```

```r
knitr::kable(cumu.logistic %>% tidy(conf.int=TRUE),format="html",digits=3)
```

y.level

term

estimate

std.error

statistic

p.value

conf.low

conf.high

cat_2

(Intercept)

0.000000e+00

11.799

-1.096000e+00

0.273

0.000000e+00

2.664217e+04

cat_2

pharmacy_num_ptt

4.310000e-01

0.867

-9.700000e-01

0.332

7.900000e-02

2.357000e+00

cat_2

most_dist_channelRETAIL PHARMACY

4.150000e-01

1.591

-5.530000e-01

0.580

1.800000e-02

9.371000e+00

cat_2

dominanceYes

1.237658e+19

1.074

4.092500e+01

0.000

1.507392e+18

1.016190e+20

cat_2

log(median_income)

5.018000e+00

1.286

1.254000e+00

0.210

4.040000e-01

6.236900e+01

cat_2

political_affRepublican

3.500000e-02

5.768

-5.800000e-01

0.562

0.000000e+00

2.857833e+03

cat_2

act_wt_person_county

2.298478e+07

10.163

1.668000e+00

0.095

5.100000e-02

1.028944e+16

cat_2

perc_oxy

9.640000e-01

0.067

-5.460000e-01

0.585

8.460000e-01

1.099000e+00

cat_2

distr_num_ptt

1.454000e+00

0.636

5.880000e-01

0.556

4.180000e-01

5.056000e+00

cat_3

(Intercept)

7.055038e+09

11.246

2.017000e+00

0.044

1.889000e+00

2.634842e+19

cat_3

pharmacy_num_ptt

6.920000e-01

0.904

-4.080000e-01

0.683

1.180000e-01

4.065000e+00

cat_3

most_dist_channelRETAIL PHARMACY

2.940000e-01

1.724

-7.110000e-01

0.477

1.000000e-02

8.616000e+00

cat_3

dominanceYes

2.607483e+19

1.003

4.457400e+01

0.000

3.651535e+18

1.861948e+20

cat_3

log(median_income)

1.640000e-01

1.264

-1.431000e+00

0.153

1.400000e-02

1.952000e+00

cat_3

political_affRepublican

9.000000e-03

5.757

-8.110000e-01

0.417

0.000000e+00

7.436080e+02

cat_3

act_wt_person_county

6.745323e+09

10.434

2.169000e+00

0.030

8.858000e+00

5.136362e+18

cat_3

perc_oxy

9.560000e-01

0.070

-6.490000e-01

0.517

8.330000e-01

1.096000e+00

cat_3

distr_num_ptt

1.135000e+00

0.658

1.920000e-01

0.848

3.120000e-01

4.122000e+00

cat_4

(Intercept)

6.997226e+36

2.342

3.622400e+01

0.000

7.101836e+34

6.894157e+38

cat_4

pharmacy_num_ptt

6.490000e-01

0.992

-4.350000e-01

0.663

9.300000e-02

4.537000e+00

cat_4

most_dist_channelRETAIL PHARMACY

4.040000e-01

1.817

-4.990000e-01

0.618

1.100000e-02

1.421500e+01

cat_4

dominanceYes

1.563295e+19

1.136

3.888800e+01

0.000

1.685258e+18

1.450159e+20

cat_4

log(median_income)

0.000000e+00

0.785

-1.045100e+01

0.000

0.000000e+00

1.000000e-03

cat_4

political_affRepublican

1.200000e-02

5.787

-7.600000e-01

0.447

0.000000e+00

1.036546e+03

cat_4

act_wt_person_county

1.336881e+11

10.644

2.407000e+00

0.016

1.165000e+02

1.534115e+20

cat_4

perc_oxy

1.021000e+00

0.076

2.680000e-01

0.789

8.790000e-01

1.184000e+00

cat_4

distr_num_ptt

1.213000e+00

0.698

2.770000e-01

0.782

3.090000e-01

4.762000e+00

cat_5

(Intercept)

1.000000e-03

0.458

-1.561900e+01

0.000

0.000000e+00

2.000000e-03

cat_5

pharmacy_num_ptt

9.085302e+13

7.286

4.411000e+00

0.000

5.707366e+07

1.446249e+20

cat_5

most_dist_channelRETAIL PHARMACY

1.636950e+02

16.608

3.070000e-01

0.759

0.000000e+00

2.243702e+16

cat_5

dominanceYes

3.185422e+44

2.803

3.655200e+01

0.000

1.308763e+42

7.753057e+46

cat_5

log(median_income)

0.000000e+00

4.087

-2.463000e+00

0.014

0.000000e+00

1.280000e-01

cat_5

political_affRepublican

1.000000e-02

7.505

-6.160000e-01

0.538

0.000000e+00

2.408523e+04

cat_5

act_wt_person_county

0.000000e+00

3.238

-5.781800e+01

0.000

0.000000e+00

0.000000e+00

cat_5

perc_oxy

2.930000e+00

0.399

2.693000e+00

0.007

1.340000e+00

6.407000e+00

cat_5

distr_num_ptt

0.000000e+00

4.466

-3.502000e+00

0.000

0.000000e+00

1.000000e-03

cat_6

(Intercept)

3.708671e+43

0.935

1.073090e+02

0.000

5.935223e+42

2.317392e+44

cat_6

pharmacy_num_ptt

4.150000e-01

1.302

-6.760000e-01

0.499

3.200000e-02

5.320000e+00

cat_6

most_dist_channelRETAIL PHARMACY

8.800000e-02

2.548

-9.540000e-01

0.340

1.000000e-03

1.298400e+01

cat_6

dominanceYes

1.259438e+20

1.854

2.495800e+01

0.000

3.324512e+18

4.771182e+21

cat_6

log(median_income)

0.000000e+00

0.873

-1.117300e+01

0.000

0.000000e+00

0.000000e+00

cat_6

political_affRepublican

0.000000e+00

0.000

-8.361746e+13

0.000

0.000000e+00

0.000000e+00

cat_6

act_wt_person_county

1.264373e+18

12.828

3.249000e+00

0.001

1.522884e+07

1.049744e+29

cat_6

perc_oxy

9.290000e-01

0.099

-7.460000e-01

0.455

7.660000e-01

1.127000e+00

cat_6

distr_num_ptt

1.662000e+00

0.868

5.860000e-01

0.558

3.030000e-01

9.106000e+00

```
# fit0.interact and fit1_interact_ord are the SAME
##########################################################################
##ordinal with interactions
fit1_interact_ord<-nnet::multinom(est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance
```

```
## # weights:  108 (85 variable)
## initial  value 191.718263
## iter  10 value 131.073412
## iter  20 value 118.096117
## iter  30 value 110.830122
## iter  40 value 98.770676
## iter  50 value 90.942099
## iter  60 value 83.536417
## iter  70 value 79.852409
## iter  80 value 76.558644
```

```
## iter  90 value 75.388227
## iter 100 value 73.332613
## final  value 73.332613
## stopped after 100 iterations
```

```r
summary(fit1_interact_ord)
```

```
## Call:
## nnet::multinom(formula = est_death_rate_cat ~ pharmacy_num_ptt +
##     most_dist_channel + dominance + log(median_income) + political_aff +
##     act_wt_person_county + perc_oxy + distr_num_ptt + log(median_income) *
##     political_aff + act_wt_person_county * distr_num_ptt + act_wt_person_county *
##     pharmacy_num_ptt + pharmacy_num_ptt * political_aff + log(median_income) *
##     most_dist_channel + log(median_income) * pharmacy_num_ptt +
##     perc_oxy * dominance + act_wt_person_county * perc_oxy, data = train_oh_wv_2012)
##
## Coefficients:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2  -70.039122         19.30373                       -89.134340
## cat_3   57.272614        -55.95872                        78.408428
## cat_4   13.254524        -19.41151                        18.793324
## cat_5    2.310263         33.15582                         4.226015
## cat_6   36.248195        112.81148                        11.483641
##       dominanceYes log(median_income) political_affRepublican
## cat_2     7.902782          19.782504              -23.925745
## cat_3    11.601608           6.686706               61.141209
## cat_4    15.459828           8.850545                4.377992
## cat_5    70.423648         -12.318535               -7.510619
## cat_6   -85.515095          17.166640               -4.394870
##       act_wt_person_county   perc_oxy distr_num_ptt
## cat_2           -46.379122 -0.1179295      2.861000
## cat_3            -6.067798 -0.1091219      2.654085
## cat_4            43.555643  0.3700651      3.690482
## cat_5           -22.441417  2.3399420    -25.874907
## cat_6             8.747659 -2.0718575      7.930411
##       log(median_income):political_affRepublican
## cat_2                                  -9.722467
## cat_3                                 -17.581315
## cat_4                                 -12.334707
## cat_5                                  -8.003669
## cat_6                                 -15.001181
##       act_wt_person_county:distr_num_ptt
## cat_2                          -12.40656
## cat_3                          -11.63255
## cat_4                          -13.87518
## cat_5                          -23.17532
## cat_6                          -23.99092
##       pharmacy_num_ptt:act_wt_person_county
## cat_2                              27.45755
## cat_3                              15.18193
## cat_4                              34.42985
## cat_5                             -54.60156
## cat_6                              59.77132
##       pharmacy_num_ptt:political_affRepublican
## cat_2                                18.065213
```

```
## cat_3                               16.777019
## cat_4                               17.387340
## cat_5                                2.975752
## cat_6                               13.994979
##       most_dist_channelRETAIL PHARMACY:log(median_income)
## cat_2                                            8.118516
## cat_3                                           -7.882542
## cat_4                                           -2.252188
## cat_5                                            1.155831
## cat_6                                           -1.501222
##       pharmacy_num_ptt:log(median_income) dominanceYes:perc_oxy
## cat_2                          -4.1193382             0.8982226
## cat_3                           3.3691941             0.8442034
## cat_4                          -0.6501612             0.7953694
## cat_5                           3.1198174             1.0870959
## cat_6                         -14.3072082             2.8972880
##       act_wt_person_county:perc_oxy
## cat_2                     0.4628329
## cat_3                     0.3988681
## cat_4                    -0.9349026
## cat_5                    -5.6523966
## cat_6                     0.6433017
##
## Std. Errors:
##       (Intercept) pharmacy_num_ptt most_dist_channelRETAIL PHARMACY
## cat_2   2.6650838        6.5507795                        0.3214459
## cat_3   3.0114917        7.9791174                        0.3058579
## cat_4   4.8929145       11.7736173                        0.5091622
## cat_5   0.6960112        0.3705614                        0.6826628
## cat_6   0.2809398        0.3182387                        0.1076812
##       dominanceYes log(median_income) political_affRepublican
## cat_2    2.9929066           1.833568            2.801574e+00
## cat_3    2.8652957           1.849142            2.923999e+00
## cat_4    4.0429297           1.868611            4.186294e+00
## cat_5    0.6825959           7.489523            5.637889e-01
## cat_6    0.7120080           2.927212            7.424363e-06
##       act_wt_person_county  perc_oxy distr_num_ptt
## cat_2           13.0786531 0.1122101     0.9950559
## cat_3           13.7169841 0.1266594     0.9734352
## cat_4            1.7099773 0.1390232     1.1445612
## cat_5            0.3470802 2.6505067     5.8107922
## cat_6            0.5262090 0.4412734     3.1089603
##       log(median_income):political_affRepublican
## cat_2                               1.904094e+00
## cat_3                               1.905924e+00
## cat_4                               1.951306e+00
## cat_5                               5.315860e+00
## cat_6                               7.856089e-05
##       act_wt_person_county:distr_num_ptt
## cat_2                           3.010011
## cat_3                           2.254038
## cat_4                           2.612740
## cat_5                           1.218282
## cat_6                           4.762232
```

```
##       pharmacy_num_ptt:act_wt_person_county
## cat_2                             3.8568923
## cat_3                             4.2597064
## cat_4                             4.3447002
## cat_5                             1.0087462
## cat_6                             0.9246094
##       pharmacy_num_ptt:political_affRepublican
## cat_2                             5.824328e+00
## cat_3                             5.853657e+00
## cat_4                             5.925844e+00
## cat_5                             4.425654e+00
## cat_6                             1.445569e-05
##       most_dist_channelRETAIL PHARMACY:log(median_income)
## cat_2                                           0.1755340
## cat_3                                           0.1952513
## cat_4                                           0.2154127
## cat_5                                           7.3046555
## cat_6                                           0.7158730
##       pharmacy_num_ptt:log(median_income) dominanceYes:perc_oxy
## cat_2                           1.0349992             0.4540956
## cat_3                           1.1998689             0.4539692
## cat_4                           1.5674393             0.4567686
## cat_5                           4.5805695             1.8063398
## cat_6                           0.5875612             0.4623058
##       act_wt_person_county:perc_oxy
## cat_2                     0.2798651
## cat_3                     0.3298717
## cat_4                     0.3010169
## cat_5                    13.6831843
## cat_6                     0.5922605
##
## Residual Deviance: 146.6652
## AIC: 316.6652
```

```
mostImportantVariables.ord.interact <- varImp(fit1_interact_ord)
mostImportantVariables.ord.interact$Variables <- row.names(mostImportantVariables.ord.interact)
mostImportantVariables.ord.interact <- mostImportantVariables.ord.interact[order(-mostImportantVariable
print(head(mostImportantVariables.ord.interact))
```

```
##                                         Overall
## pharmacy_num_ptt                       240.6413
## most_dist_channelRETAIL PHARMACY       202.0458
## pharmacy_num_ptt:act_wt_person_county 191.4422
## dominanceYes                           190.9030
## act_wt_person_county                   127.1916
## political_affRepublican                101.3504
##                                                                   Variables
## pharmacy_num_ptt                                           pharmacy_num_ptt
## most_dist_channelRETAIL PHARMACY           most_dist_channelRETAIL PHARMACY
## pharmacy_num_ptt:act_wt_person_county pharmacy_num_ptt:act_wt_person_county
## dominanceYes                                                   dominanceYes
## act_wt_person_county                                   act_wt_person_county
## political_affRepublican                             political_affRepublican
```

```r
knitr::kable(fit1_interact_ord %>% tidy(conf.int=TRUE),format="html",digits=3)
```

y.level

term

estimate

std.error

statistic

p.value

conf.low

conf.high

cat_2

(Intercept)

0.000000e+00

2.665

-26.280

0.000

0.000000e+00

0.000000e+00

cat_2

pharmacy_num_ptt

2.418252e+08

6.551

2.947

0.003

6.419010e+02

9.110353e+13

cat_2

most_dist_channelRETAIL PHARMACY

0.000000e+00

0.321

-277.292

0.000

0.000000e+00

0.000000e+00

cat_2

dominanceYes

2.704797e+03

2.993

2.641

0.008

7.666000e+00

9.543362e+05

cat_2

log(median_income)

3.903302e+08

1.834

10.789

0.000

1.073239e+07

1.419607e+10

cat_2

political_affRepublican

0.000000e+00

2.802

-8.540

0.000

0.000000e+00

0.000000e+00

cat_2

act_wt_person_county

0.000000e+00

13.079

-3.546

0.000

0.000000e+00

0.000000e+00

cat_2

perc_oxy

8.890000e-01

0.112

-1.051

0.293

7.130000e-01

1.107000e+00

cat_2

distr_num_ptt

1.747900e+01

0.995

2.875

0.004

2.486000e+00

1.228880e+02

cat_2

log(median_income):political_affRepublican

0.000000e+00

1.904

-5.106

0.000

0.000000e+00

3.000000e-03

cat_2

act_wt_person_county:distr_num_ptt

0.000000e+00

3.010

-4.122

0.000

0.000000e+00

1.000000e-03

cat_2

pharmacy_num_ptt:act_wt_person_county

8.407374e+11

3.857

7.119

0.000

4.381952e+08

1.613069e+15

cat_2

pharmacy_num_ptt:political_affRepublican

7.008456e+07

5.824

3.102

0.002

7.725830e+02

6.357689e+12

cat_2

most_dist_channelRETAIL PHARMACY:log(median_income)

3.356037e+03

0.176

46.250

0.000

2.379096e+03

4.734145e+03

cat_2

pharmacy_num_ptt:log(median_income)

1.600000e-02

1.035

-3.980

0.000

2.000000e-03

1.240000e-01

cat_2

dominanceYes:perc_oxy

2.455000e+00

0.454

1.978

0.048

1.008000e+00

5.979000e+00

cat_2

act_wt_person_county:perc_oxy

1.589000e+00

0.280

1.654

0.098

9.180000e-01

2.749000e+00

cat_3

(Intercept)

7.467588e+24

3.011

19.018

0.000

2.040770e+22

2.732541e+27

cat_3

pharmacy_num_ptt

0.000000e+00

7.979

-7.013

0.000

0.000000e+00

0.000000e+00

cat_3

most_dist_channelRETAIL PHARMACY

1.128100e+34

0.306

256.356

0.000

6.194425e+33

2.054445e+34

cat_3

dominanceYes

1.092734e+05

2.865

4.049

0.000

3.977140e+02

3.002332e+07

cat_3

log(median_income)

8.016770e+02

1.849

3.616

0.000

2.138000e+01

3.006019e+04

cat_3

political_affRepublican

3.575113e+26

2.924

20.910

0.000

1.159785e+24

1.102053e+29

cat_3

act_wt_person_county

2.000000e-03

13.717

-0.442

0.658

0.000000e+00

1.098264e+09

cat_3

perc_oxy

8.970000e-01

0.127

-0.862

0.389

7.000000e-01

1.149000e+00

cat_3

distr_num_ptt

1.421200e+01

0.973

2.727

0.006

2.109000e+00

9.577300e+01

cat_3

log(median_income):political_affRepublican

0.000000e+00

1.906

-9.225

0.000

0.000000e+00

0.000000e+00

cat_3

act_wt_person_county:distr_num_ptt

0.000000e+00

2.254

-5.161

0.000

0.000000e+00

1.000000e-03

cat_3

pharmacy_num_ptt:act_wt_person_county

3.921274e+06

4.260

3.564

0.000

9.280230e+02

1.656898e+10

cat_3

pharmacy_num_ptt:political_affRepublican

1.932711e+07

5.854

2.866

0.004

2.011520e+02

1.856988e+12

cat_3

most_dist_channelRETAIL PHARMACY:log(median_income)

0.000000e+00

0.195

-40.371

0.000

0.000000e+00

1.000000e-03

cat_3

pharmacy_num_ptt:log(median_income)

2.905500e+01

1.200

2.808

0.005

2.766000e+00

3.051780e+02

cat_3

dominanceYes:perc_oxy

2.326000e+00

0.454

1.860

0.063

9.550000e-01

5.663000e+00

cat_3

act_wt_person_county:perc_oxy

1.490000e+00

0.330

1.209

0.227

7.810000e-01

2.845000e+00

cat_4

(Intercept)

5.706457e+05

4.893

2.709

0.007

3.904000e+01

8.341092e+09

cat_4

pharmacy_num_ptt

0.000000e+00

11.774

-1.649

0.099

0.000000e+00

3.903100e+01

cat_4

most_dist_channelRETAIL PHARMACY

1.451567e+08

0.509

36.910

0.000

5.351031e+07

3.937647e+08

cat_4

dominanceYes

5.177477e+06

4.043

3.824

0.000

1.874002e+03

1.430429e+10

cat_4

log(median_income)

6.978192e+03

1.869

4.736

0.000

1.791340e+02

2.718361e+05

cat_4

political_affRepublican

7.967800e+01

4.186

1.046

0.296

2.200000e-02

2.915530e+05

cat_4

act_wt_person_county

8.240913e+18

1.710

25.471

0.000

2.886955e+17

2.352397e+20

cat_4

perc_oxy

1.448000e+00

0.139

2.662

0.008

1.103000e+00

1.901000e+00

cat_4

distr_num_ptt

4.006400e+01

1.145

3.224

0.001

4.251000e+00

3.775790e+02

cat_4

log(median_income):political_affRepublican

0.000000e+00

1.951

-6.321

0.000

0.000000e+00

0.000000e+00

cat_4

act_wt_person_county:distr_num_ptt

0.000000e+00

2.613

-5.311

0.000

0.000000e+00

0.000000e+00

cat_4

pharmacy_num_ptt:act_wt_person_county

8.967960e+14

4.345

7.925

0.000

1.796711e+11

4.476197e+18

cat_4

pharmacy_num_ptt:political_affRepublican

3.558162e+07

5.926

2.934

0.003

3.214680e+02

3.938343e+12

cat_4

most_dist_channelRETAIL PHARMACY:log(median_income)

1.050000e-01

0.215

-10.455

0.000

6.900000e-02

1.600000e-01

cat_4

pharmacy_num_ptt:log(median_income)

5.220000e-01

1.567

-0.415

0.678

2.400000e-02

1.126800e+01

cat_4

dominanceYes:perc_oxy

2.215000e+00

0.457

1.741

0.082

9.050000e-01

5.423000e+00

cat_4

act_wt_person_county:perc_oxy

3.930000e-01

0.301

-3.106

0.002

2.180000e-01

7.080000e-01

cat_5

(Intercept)

1.007700e+01

0.696

3.319

0.001

2.576000e+00

3.942600e+01

cat_5

pharmacy_num_ptt

2.508359e+14

0.371

89.475

0.000

1.213298e+14

5.185757e+14

cat_5

most_dist_channelRETAIL PHARMACY

6.844400e+01

0.683

6.190

0.000

1.795800e+01

2.608670e+02

cat_5

dominanceYes

3.842392e+30

0.683

103.170

0.000

1.008266e+30

1.464294e+31

cat_5

log(median_income)

0.000000e+00

7.490

-1.645

0.100

0.000000e+00

1.059800e+01

cat_5

political_affRepublican

1.000000e-03

0.564

-13.322

0.000

0.000000e+00

2.000000e-03

cat_5

act_wt_person_county

0.000000e+00

0.347

-64.658

0.000

0.000000e+00

0.000000e+00

cat_5

perc_oxy

1.038100e+01

2.651

0.883

0.377

5.800000e-02

1.872144e+03

cat_5

distr_num_ptt

0.000000e+00

5.811

-4.453

0.000

0.000000e+00

0.000000e+00

cat_5

log(median_income):political_affRepublican

0.000000e+00

5.316

-1.506

0.132

0.000000e+00

1.119200e+01

cat_5

act_wt_person_county:distr_num_ptt

0.000000e+00

1.218

-19.023

0.000

0.000000e+00

0.000000e+00

cat_5

pharmacy_num_ptt:act_wt_person_county

0.000000e+00

1.009

-54.128

0.000

0.000000e+00

0.000000e+00

cat_5

pharmacy_num_ptt:political_affRepublican

1.960400e+01

4.426

0.672

0.501

3.000000e-03

1.146770e+05

cat_5

most_dist_channelRETAIL PHARMACY:log(median_income)

3.177000e+00

7.305

0.158

0.874

0.000000e+00

5.244509e+06

cat_5

pharmacy_num_ptt:log(median_income)

2.264200e+01

4.581

0.681

0.496

3.000000e-03

1.794351e+05

cat_5

dominanceYes:perc_oxy

2.966000e+00

1.806

0.602

0.547

8.600000e-02

1.022540e+02

cat_5

act_wt_person_county:perc_oxy

4.000000e-03

13.683

-0.413

0.680

0.000000e+00

1.557195e+09

cat_6

(Intercept)

5.525749e+15

0.281

129.025

0.000

3.186066e+15

9.583574e+15

cat_6

pharmacy_num_ptt

9.849221e+48

0.318

354.487

0.000

5.278574e+48

1.837753e+49

cat_6

most_dist_channelRETAIL PHARMACY

9.711406e+04

0.108

106.645

0.000

7.863641e+04

1.199335e+05

cat_6

dominanceYes

0.000000e+00

0.712

-120.104

0.000

0.000000e+00

0.000000e+00

cat_6

log(median_income)

2.853495e+07

2.927

5.865

0.000

9.198771e+04

8.851655e+09

cat_6

political_affRepublican

1.200000e-02

0.000

-591952.396

0.000

1.200000e-02

1.200000e-02

cat_6

act_wt_person_county

6.295935e+03

0.526

16.624

0.000

2.244659e+03

1.765916e+04

cat_6

perc_oxy

1.260000e-01

0.441

-4.695

0.000

5.300000e-02

2.990000e-01

cat_6

distr_num_ptt

2.780570e+03

3.109

2.551

0.011

6.277000e+00

1.231646e+06

cat_6

log(median_income):political_affRepublican

0.000000e+00

0.000

-190949.727

0.000

0.000000e+00

0.000000e+00

cat_6

act_wt_person_county:distr_num_ptt

0.000000e+00

4.762

-5.038

0.000

0.000000e+00

0.000000e+00

cat_6

pharmacy_num_ptt:act_wt_person_county

9.085618e+25

0.925

64.645

0.000

1.483630e+25

5.563953e+26

cat_6

pharmacy_num_ptt:political_affRepublican

1.196581e+06

0.000

968129.680

0.000

1.196547e+06

1.196615e+06

cat_6

most_dist_channelRETAIL PHARMACY:log(median_income)

2.230000e-01

0.716

-2.097

0.036

5.500000e-02

9.070000e-01

cat_6

pharmacy_num_ptt:log(median_income)

0.000000e+00

0.588

-24.350

0.000

0.000000e+00

0.000000e+00

cat_6

dominanceYes:perc_oxy

1.812500e+01

0.462

6.267

0.000

7.324000e+00

4.485300e+01

cat_6

act_wt_person_county:perc_oxy

1.903000e+00

0.592

1.086

0.277

5.960000e-01

6.075000e+00

Cumulative logistic regression

```
### polyr
library(MASS)
fit1<-polr(est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance + log(median_income) +
summary(fit1)
```

```
## Call:
## polr(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     perc_oxy + distr_num_ptt, data = train_oh_wv_2012, Hess = TRUE,
##     method = "logistic")
##
## Coefficients:
##                                      Value Std. Error t value
## pharmacy_num_ptt                  0.564486     0.2747  2.0549
## most_dist_channelRETAIL PHARMACY -0.254243     0.4870 -0.5220
## dominanceYes                      1.547301     0.6263  2.4706
## log(median_income)               -4.415956     1.3485 -3.2747
## political_affRepublican          -1.304081     0.4717 -2.7644
## act_wt_person_county              6.993948     1.9732  3.5445
## perc_oxy                          0.003462     0.0203  0.1706
## distr_num_ptt                    -0.275981     0.1647 -1.6759
##
## Intercepts:
##             Value    Std. Error t value
## cat_1|cat_2 -48.5503  14.6190    -3.3210
## cat_2|cat_3 -44.9383  14.4714    -3.1053
## cat_3|cat_4 -43.0232  14.4176    -2.9841
## cat_4|cat_5 -41.2075  14.3968    -2.8623
## cat_5|cat_6 -40.1057  14.4018    -2.7848
##
## Residual Deviance: 244.838
## AIC: 270.838
```

```
ctable <- coef(summary(fit1))
## calculate and store p values
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
## combined table
(ctable <- cbind(ctable, "p value" = p))
```

```
##                                          Value  Std. Error    t value
## pharmacy_num_ptt                   0.564486010  0.27470434  2.0548856
## most_dist_channelRETAIL PHARMACY  -0.254243035  0.48703105 -0.5220263
## dominanceYes                       1.547301283  0.62629680  2.4705559
## log(median_income)                -4.415955555  1.34850170 -3.2747126
```

```
## political_affRepublican           -1.304081141  0.47173549 -2.7644329
## act_wt_person_county               6.993947655  1.97317983  3.5445060
## perc_oxy                           0.003462341  0.02029621  0.1705905
## distr_num_ptt                     -0.275981147  0.16467758 -1.6758878
## cat_1|cat_2                      -48.550313642 14.61901106 -3.3210395
## cat_2|cat_3                      -44.938314787 14.47136335 -3.1053270
## cat_3|cat_4                      -43.023163517 14.41762188 -2.9840680
## cat_4|cat_5                      -41.207454790 14.39684221 -2.8622565
## cat_5|cat_6                      -40.105733160 14.40175107 -2.7847817
##                                      p value
## pharmacy_num_ptt                 0.0398900519
## most_dist_channelRETAIL PHARMACY 0.6016519908
## dominanceYes                     0.0134903220
## log(median_income)               0.0010576950
## political_affRepublican          0.0057021825
## act_wt_person_county             0.0003933496
## perc_oxy                         0.8645457567
## distr_num_ptt                    0.0937601746
## cat_1|cat_2                      0.0008968283
## cat_2|cat_3                      0.0019006879
## cat_3|cat_4                      0.0028444359
## cat_4|cat_5                      0.0042063637
## cat_5|cat_6                      0.0053563767
```

```
#not significant at p=0.05: most dist channel retail, perc_oxy
```

```
### polyr with interactions
fit1_interact<-polr(est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance + log(median_i
summary(fit1_interact)
```

```
## Call:
## polr(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     perc_oxy + distr_num_ptt + log(median_income) * political_aff +
##     act_wt_person_county * distr_num_ptt + act_wt_person_county *
##     pharmacy_num_ptt + pharmacy_num_ptt * political_aff + log(median_income) *
##     most_dist_channel + log(median_income) * pharmacy_num_ptt +
##     perc_oxy * dominance + act_wt_person_county * perc_oxy, data = train_oh_wv_2012,
##     Hess = TRUE, method = "logistic")
##
## Coefficients:
##                                                         Value Std. Error
## pharmacy_num_ptt                                    -28.11061    7.60427
## most_dist_channelRETAIL PHARMACY                     55.47383    5.93747
## dominanceYes                                          0.35146    2.46953
## log(median_income)                                  -13.17296    0.70656
## political_affRepublican                             -34.69906   17.45942
## act_wt_person_county                                 15.68627   12.15334
## perc_oxy                                              0.02317    0.05583
## distr_num_ptt                                        -0.20047    0.38816
## log(median_income):political_affRepublican            3.32787    1.60366
## act_wt_person_county:distr_num_ptt                    0.11971    1.25706
## pharmacy_num_ptt:act_wt_person_county                 0.30237    1.69157
## pharmacy_num_ptt:political_affRepublican             -0.76800    0.49133
## most_dist_channelRETAIL PHARMACY:log(median_income)  -5.28874    0.56043
```

```
## pharmacy_num_ptt:log(median_income)                       2.72185    0.74533
## dominanceYes:perc_oxy                                      0.01549    0.03995
## act_wt_person_county:perc_oxy                             -0.14413    0.17242
##                                                            t value
## pharmacy_num_ptt                                          -3.69669
## most_dist_channelRETAIL PHARMACY                           9.34300
## dominanceYes                                               0.14232
## log(median_income)                                       -18.64386
## political_affRepublican                                   -1.98741
## act_wt_person_county                                       1.29070
## perc_oxy                                                   0.41503
## distr_num_ptt                                             -0.51647
## log(median_income):political_affRepublican                 2.07518
## act_wt_person_county:distr_num_ptt                         0.09523
## pharmacy_num_ptt:act_wt_person_county                      0.17875
## pharmacy_num_ptt:political_affRepublican                  -1.56311
## most_dist_channelRETAIL PHARMACY:log(median_income)       -9.43697
## pharmacy_num_ptt:log(median_income)                        3.65189
## dominanceYes:perc_oxy                                      0.38784
## act_wt_person_county:perc_oxy                             -0.83593
##
## Intercepts:
##             Value     Std. Error t value
## cat_1|cat_2 -139.4913    6.3045   -22.1256
## cat_2|cat_3 -135.7621    6.2292   -21.7943
## cat_3|cat_4 -133.6689    6.2160   -21.5038
## cat_4|cat_5 -131.6033    6.2678   -20.9968
## cat_5|cat_6 -130.4057    6.3071   -20.6760
##
## Residual Deviance: 232.0324
## AIC: 274.0324
```

```r
ctable.interact.1 <- coef(summary(fit1_interact))
## calculate and store p values
p1.interact <- pnorm(abs(ctable.interact.1[, "t value"]), lower.tail = FALSE) * 2
## combined table
(ctable.interact.1 <- cbind(ctable.interact.1, "p value" = p1.interact))
```

```
##                                                                Value
## pharmacy_num_ptt                                          -28.11061084
## most_dist_channelRETAIL PHARMACY                           55.47383338
## dominanceYes                                                0.35146064
## log(median_income)                                        -13.17296462
## political_affRepublican                                   -34.69906090
## act_wt_person_county                                       15.68626835
## perc_oxy                                                    0.02317058
## distr_num_ptt                                              -0.20047156
## log(median_income):political_affRepublican                  3.32786901
## act_wt_person_county:distr_num_ptt                          0.11970531
## pharmacy_num_ptt:act_wt_person_county                       0.30237478
## pharmacy_num_ptt:political_affRepublican                   -0.76799748
## most_dist_channelRETAIL PHARMACY:log(median_income)        -5.28874487
## pharmacy_num_ptt:log(median_income)                         2.72185016
## dominanceYes:perc_oxy                                       0.01549378
## act_wt_person_county:perc_oxy                              -0.14412717
```

```
## cat_1|cat_2                                      -139.49126592
## cat_2|cat_3                                      -135.76205566
## cat_3|cat_4                                      -133.66889301
## cat_4|cat_5                                      -131.60331423
## cat_5|cat_6                                      -130.40574916
##                                                    Std. Error
## pharmacy_num_ptt                                   7.60427366
## most_dist_channelRETAIL PHARMACY                   5.93747255
## dominanceYes                                       2.46953118
## log(median_income)                                 0.70655771
## political_affRepublican                           17.45942266
## act_wt_person_county                              12.15333911
## perc_oxy                                           0.05582871
## distr_num_ptt                                      0.38815727
## log(median_income):political_affRepublican         1.60365661
## act_wt_person_county:distr_num_ptt                 1.25706330
## pharmacy_num_ptt:act_wt_person_county              1.69157374
## pharmacy_num_ptt:political_affRepublican           0.49132750
## most_dist_channelRETAIL PHARMACY:log(median_income)  0.56042836
## pharmacy_num_ptt:log(median_income)                0.74532583
## dominanceYes:perc_oxy                              0.03994851
## act_wt_person_county:perc_oxy                      0.17241553
## cat_1|cat_2                                        6.30450681
## cat_2|cat_3                                        6.22924177
## cat_3|cat_4                                        6.21604690
## cat_4|cat_5                                        6.26777883
## cat_5|cat_6                                        6.30710783
##                                                        t value
## pharmacy_num_ptt                                   -3.69668585
## most_dist_channelRETAIL PHARMACY                    9.34300461
## dominanceYes                                        0.14231877
## log(median_income)                                -18.64386224
## political_affRepublican                            -1.98741170
## act_wt_person_county                                1.29069618
## perc_oxy                                            0.41502988
## distr_num_ptt                                      -0.51646994
## log(median_income):political_affRepublican          2.07517557
## act_wt_person_county:distr_num_ptt                  0.09522616
## pharmacy_num_ptt:act_wt_person_county               0.17875353
## pharmacy_num_ptt:political_affRepublican           -1.56310705
## most_dist_channelRETAIL PHARMACY:log(median_income)  -9.43696874
## pharmacy_num_ptt:log(median_income)                 3.65189296
## dominanceYes:perc_oxy                               0.38784389
## act_wt_person_county:perc_oxy                      -0.83592916
## cat_1|cat_2                                       -22.12564282
## cat_2|cat_3                                       -21.79431474
## cat_3|cat_4                                       -21.50384243
## cat_4|cat_5                                       -20.99680251
## cat_5|cat_6                                       -20.67599805
##                                                         p value
## pharmacy_num_ptt                                   2.184324e-04
## most_dist_channelRETAIL PHARMACY                   9.363838e-21
## dominanceYes                                       8.868282e-01
## log(median_income)                                 1.416357e-77
```

```
## political_affRepublican                              4.687679e-02
## act_wt_person_county                                 1.968091e-01
## perc_oxy                                             6.781200e-01
## distr_num_ptt                                        6.055262e-01
## log(median_income):political_affRepublican           3.797027e-02
## act_wt_person_county:distr_num_ptt                   9.241352e-01
## pharmacy_num_ptt:act_wt_person_county                8.581312e-01
## pharmacy_num_ptt:political_affRepublican             1.180274e-01
## most_dist_channelRETAIL PHARMACY:log(median_income)  3.837194e-21
## pharmacy_num_ptt:log(median_income)                  2.603144e-04
## dominanceYes:perc_oxy                                6.981316e-01
## act_wt_person_county:perc_oxy                        4.031948e-01
## cat_1|cat_2                                         1.790677e-108
## cat_2|cat_3                                         2.626947e-105
## cat_3|cat_4                                         1.433166e-102
## cat_4|cat_5                                          7.015095e-98
## cat_5|cat_6                                          5.697604e-95
```

Backward selection to get lowest AIC (cumulative logit with interactions) - FINAL MODEL

```
## fit significant predictors and interations with cumulative logit
fit.select <- stepAIC(fit1_interact, trace = FALSE)
summary(fit.select)
```

```
## Call:
## polr(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     pharmacy_num_ptt:political_aff + most_dist_channel:log(median_income) +
##     pharmacy_num_ptt:log(median_income), data = train_oh_wv_2012,
##     Hess = TRUE, method = "logistic")
##
## Coefficients:
##                                                        Value Std. Error
## pharmacy_num_ptt                                    -29.8071    13.1492
## most_dist_channelRETAIL PHARMACY                     66.0704    27.7312
## dominanceYes                                          0.9971     0.5648
## log(median_income)                                   -9.9542     3.4002
## political_affRepublican                               1.0636     1.3325
## act_wt_person_county                                  7.5644     1.7395
## pharmacy_num_ptt:political_affRepublican             -0.9046     0.4408
## most_dist_channelRETAIL PHARMACY:log(median_income)  -6.2989     2.6236
## pharmacy_num_ptt:log(median_income)                   2.8877     1.2485
##                                                     t value
## pharmacy_num_ptt                                    -2.2668
## most_dist_channelRETAIL PHARMACY                     2.3825
## dominanceYes                                         1.7653
## log(median_income)                                  -2.9275
## political_affRepublican                              0.7982
## act_wt_person_county                                 4.3485
## pharmacy_num_ptt:political_affRepublican            -2.0523
## most_dist_channelRETAIL PHARMACY:log(median_income) -2.4009
## pharmacy_num_ptt:log(median_income)                  2.3129
##
```

```
## Intercepts:
##             Value     Std. Error t value
## cat_1|cat_2 -105.9514   36.0762    -2.9369
## cat_2|cat_3 -102.2201   35.9160    -2.8461
## cat_3|cat_4 -100.1991   35.8594    -2.7942
## cat_4|cat_5  -98.2888   35.8476    -2.7419
## cat_5|cat_6  -97.1360   35.8385    -2.7104
##
## Residual Deviance: 236.3985
## AIC: 264.3985
```

```
ctable.interact.2 <- coef(summary(fit.select))
## calculate and store p values
p2.interact <- pnorm(abs(ctable.interact.2[, "t value"]), lower.tail = FALSE) * 2
(ctable.interact.2 <- cbind(ctable.interact.2, "p value" = p2.interact))
```

```
##                                                          Value
## pharmacy_num_ptt                                   -29.8070923
## most_dist_channelRETAIL PHARMACY                    66.0703922
## dominanceYes                                         0.9971020
## log(median_income)                                  -9.9541669
## political_affRepublican                              1.0636176
## act_wt_person_county                                 7.5643963
## pharmacy_num_ptt:political_affRepublican            -0.9046001
## most_dist_channelRETAIL PHARMACY:log(median_income)  -6.2989195
## pharmacy_num_ptt:log(median_income)                  2.8876761
## cat_1|cat_2                                       -105.9514019
## cat_2|cat_3                                       -102.2200916
## cat_3|cat_4                                       -100.1991438
## cat_4|cat_5                                        -98.2887600
## cat_5|cat_6                                        -97.1360291
##                                                    Std. Error    t value
## pharmacy_num_ptt                                   13.1491667 -2.2668427
## most_dist_channelRETAIL PHARMACY                   27.7311977  2.3825293
## dominanceYes                                        0.5648187  1.7653489
## log(median_income)                                  3.4002086 -2.9275166
## political_affRepublican                             1.3325295  0.7981944
## act_wt_person_county                                1.7395230  4.3485462
## pharmacy_num_ptt:political_affRepublican            0.4407761 -2.0522894
## most_dist_channelRETAIL PHARMACY:log(median_income)  2.6235931 -2.4008752
## pharmacy_num_ptt:log(median_income)                 1.2484834  2.3129472
## cat_1|cat_2                                        36.0761506 -2.9368821
## cat_2|cat_3                                        35.9160204 -2.8460862
## cat_3|cat_4                                        35.8594230 -2.7942207
## cat_4|cat_5                                        35.8475514 -2.7418542
## cat_5|cat_6                                        35.8385116 -2.7103812
##                                                         p value
## pharmacy_num_ptt                                   2.339984e-02
## most_dist_channelRETAIL PHARMACY                   1.719416e-02
## dominanceYes                                       7.750515e-02
## log(median_income)                                 3.416808e-03
## political_affRepublican                            4.247577e-01
## act_wt_person_county                               1.370429e-05
## pharmacy_num_ptt:political_affRepublican           4.014154e-02
## most_dist_channelRETAIL PHARMACY:log(median_income) 1.635591e-02
```

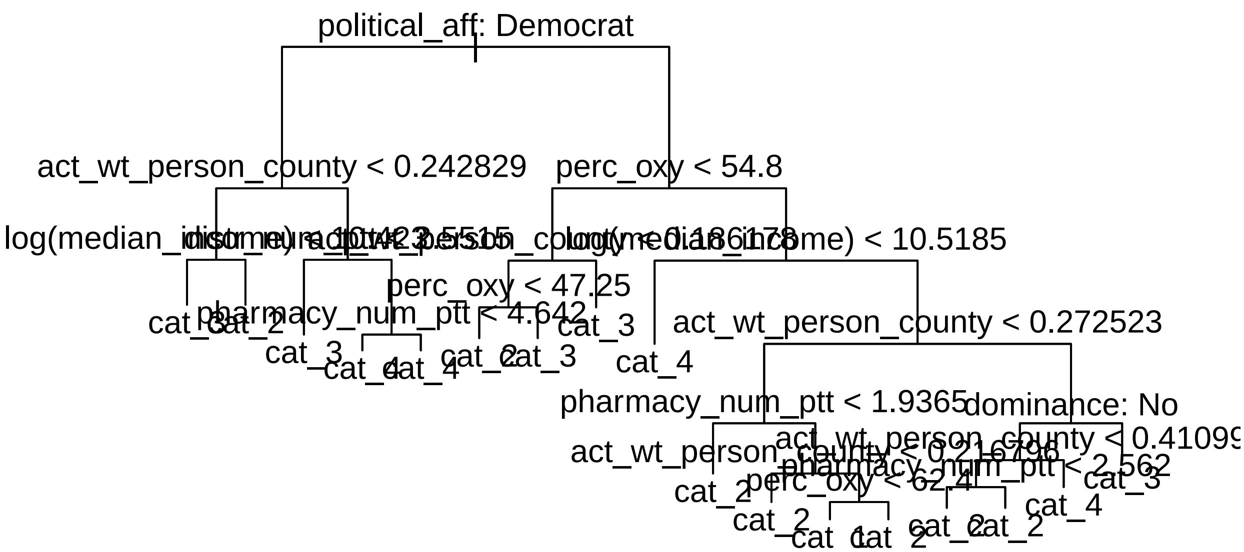
```
## pharmacy_num_ptt:log(median_income)                    2.072554e-02
## cat_1|cat_2                                             3.315301e-03
## cat_2|cat_3                                             4.426021e-03
## cat_3|cat_4                                             5.202495e-03
## cat_4|cat_5                                             6.109345e-03
## cat_5|cat_6                                             6.720591e-03
```

Tree Models

```
pred_matrix <- train_oh_wv_2012 %>%  # a matrix of predictors
  mutate(log_income = log(median_income)) %>%
  dplyr::select(est_death_rate_cat, pharmacy_num_ptt, most_dist_channel, dominance, log_income,  politi
## Classification tree model
set.seed(1)
classtree <- tree(est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance + log(median_in
summary(classtree)
```

```
##
## Classification tree:
## tree(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     perc_oxy + distr_num_ptt, data = train_oh_wv_2012)
## Variables actually used in tree construction:
## [1] "political_aff"        "act_wt_person_county" "log(median_income)"
## [4] "distr_num_ptt"        "pharmacy_num_ptt"     "perc_oxy"
## [7] "dominance"
## Number of terminal nodes:  17
## Residual mean deviance:  1.427 = 128.4 / 90
## Misclassification error rate: 0.3084 = 33 / 107
```

```
plot(classtree)
text(classtree, pretty = 0) # The most important predictor is pharmacy-num-ptt
```




```
deathrate.test <- test_oh_wv_2012$est_death_rate_cat
classtree.pred <- predict(classtree, test_oh_wv_2012, type = "class")
table(classtree.pred, deathrate.test)
```

```
##                deathrate.test
## classtree.pred cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##          cat_1     0     1     0     0     0     0
##          cat_2     3    13     2     0     0     0
##          cat_3     0     1     3     1     1     1
##          cat_4     0     1     3     0     2     2
##          cat_5     0     0     0     0     0     0
##          cat_6     0     0     0     0     0     2
```

```r
sum(diag(table(classtree.pred, deathrate.test)))/36 # correctly classified ~36%.
```

```
## [1] 0.5
```

```r
library(e1071)
```

```
##
## Attaching package: 'e1071'

## The following object is masked from 'package:brms':
##
##     rwiener
```

```r
caret::confusionMatrix(classtree.pred, deathrate.test)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##      cat_1     0     1     0     0     0     0
##      cat_2     3    13     2     0     0     0
##      cat_3     0     1     3     1     1     1
##      cat_4     0     1     3     0     2     2
##      cat_5     0     0     0     0     0     0
##      cat_6     0     0     0     0     0     2
##
## Overall Statistics
##
##                Accuracy : 0.5
##                  95% CI : (0.3292, 0.6708)
##     No Information Rate : 0.4444
##     P-Value [Acc > NIR] : 0.3061
##
##                   Kappa : 0.304
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: cat_1 Class: cat_2 Class: cat_3 Class: cat_4
## Sensitivity               0.00000       0.8125      0.37500      0.00000
## Specificity               0.96970       0.7500      0.85714      0.77143
## Pos Pred Value            0.00000       0.7222      0.42857      0.00000
## Neg Pred Value            0.91429       0.8333      0.82759      0.96429
## Prevalence                0.08333       0.4444      0.22222      0.02778
## Detection Rate            0.00000       0.3611      0.08333      0.00000
## Detection Prevalence      0.02778       0.5000      0.19444      0.22222
## Balanced Accuracy         0.48485       0.7812      0.61607      0.38571
```

```
##                      Class: cat_5 Class: cat_6
## Sensitivity               0.00000      0.40000
## Specificity               1.00000      1.00000
## Pos Pred Value                NaN      1.00000
## Neg Pred Value            0.91667      0.91176
## Prevalence                0.08333      0.13889
## Detection Rate            0.00000      0.05556
## Detection Prevalence      0.00000      0.05556
## Balanced Accuracy         0.50000      0.70000
```

```r
library(mltest)
ml_test(classtree.pred, deathrate.test, output.as.table = FALSE)
```

```
## $accuracy
## [1] 0.5
##
## $balanced.accuracy
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 0.4736842 0.6562500 0.5822368 0.3461538 0.5000000 0.7000000
##
## $DOR
##    cat_1    cat_2    cat_3    cat_4    cat_5    cat_6
## 0.000000 4.333333 2.250000 0.000000      NaN      Inf
##
## $error.rate
## [1] 0.5
##
## $F0.5
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
##       NaN 0.7386364 0.4166667       NaN       NaN 0.7692308
##
## $F1
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
##       NaN 0.7647059 0.4000000       NaN       NaN 0.5714286
##
## $F2
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
##       NaN 0.7926829 0.3846154       NaN       NaN 0.4545455
##
## $FDR
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 1.0000000 0.2777778 0.5714286 1.0000000       NaN 0.0000000
##
## $FNR
##  cat_1  cat_2  cat_3  cat_4  cat_5  cat_6
## 1.0000 0.1875 0.6250 1.0000 1.0000 0.6000
##
## $FOR
##      cat_1      cat_2      cat_3      cat_4      cat_5      cat_6
## 0.14285714 0.37500000 0.25000000 0.05263158 0.14285714 0.15789474
##
## $FPR
##      cat_1      cat_2      cat_3      cat_4      cat_5      cat_6
## 0.05263158 0.50000000 0.21052632 0.30769231 0.00000000 0.00000000
##
```

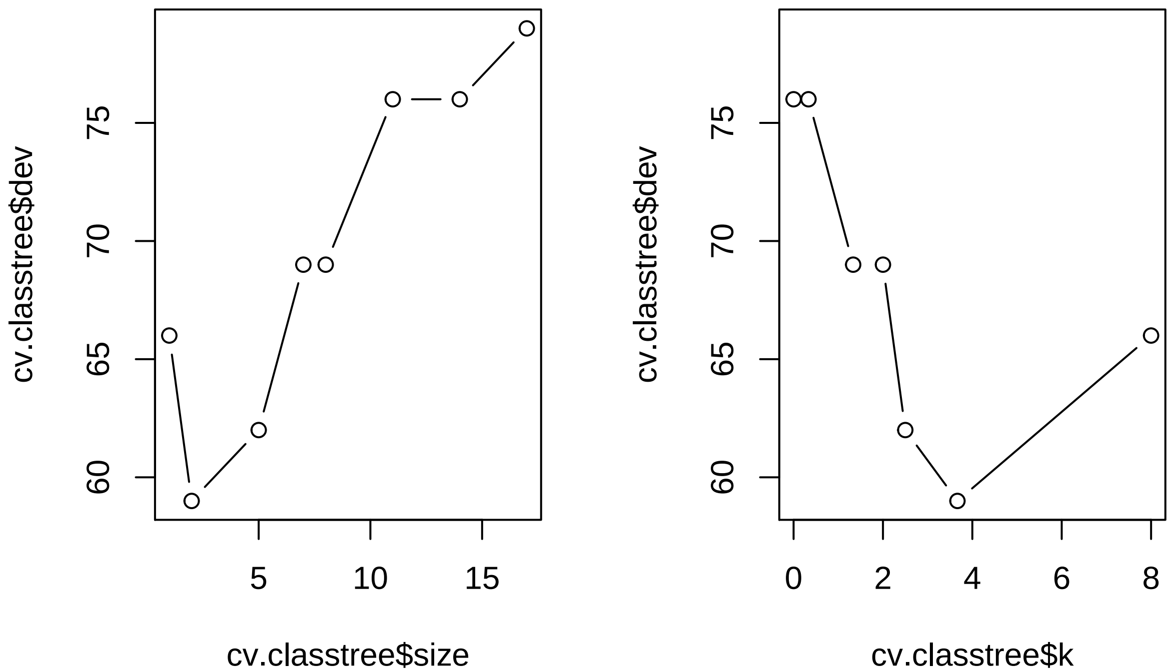
```
## $geometric.mean
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 0.0000000 0.6373774 0.5441072 0.0000000 0.0000000 0.6324555
##
## $Jaccard
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 0.0000000 0.6190476 0.2500000 0.0000000 0.0000000 0.4000000
##
## $L
##   cat_1   cat_2   cat_3   cat_4   cat_5   cat_6
## 0.00000 1.62500 1.78125 0.00000     NaN     Inf
##
## $lambda
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 1.0555556 0.3750000 0.7916667 1.4444444 1.0000000 0.6000000
##
## $MCC
##      cat_1      cat_2      cat_3      cat_4      cat_5      cat_6
## -0.0867110  0.3294039  0.1713777 -0.1272570        NaN  0.5803810
##
## $MK
##       cat_1       cat_2       cat_3       cat_4       cat_5       cat_6
## -0.14285714  0.34722222  0.17857143 -0.05263158         NaN  0.84210526
##
## $NPV
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 0.8571429 0.6250000 0.7500000 0.9473684 0.8571429 0.8421053
##
## $OP
##       cat_1       cat_2       cat_3       cat_4       cat_5       cat_6
## -0.50000000  0.26190476  0.14406780 -0.50000000 -0.50000000  0.07142857
##
## $precision
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 0.0000000 0.7222222 0.4285714 0.0000000       NaN 1.0000000
##
## $recall
##  cat_1  cat_2  cat_3  cat_4  cat_5  cat_6
## 0.0000 0.8125 0.3750 0.0000 0.0000 0.4000
##
## $specificity
##     cat_1     cat_2     cat_3     cat_4     cat_5     cat_6
## 0.9473684 0.5000000 0.7894737 0.6923077 1.0000000 1.0000000
##
## $Youden
##       cat_1       cat_2       cat_3       cat_4       cat_5       cat_6
## -0.05263158  0.31250000  0.16447368 -0.30769231  0.00000000  0.40000000
```

```
## Pruned classfiction tree model
set.seed(3)
cv.classtree <- cv.tree(classtree, FUN = prune.misclass)
cv.classtree
```

```
## $size
## [1] 17 14 11  8  7  5  2  1
```

```
##
## $dev
## [1] 79 76 76 69 69 62 59 66
##
## $k
## [1]      -Inf 0.0000000 0.3333333 1.3333333 2.0000000 2.5000000 3.6666667
## [8] 8.0000000
##
## $method
## [1] "misclass"
##
## attr(,"class")
## [1] "prune"         "tree.sequence"
```

```
par(mfrow = c(1,2))
plot(cv.classtree$size, cv.classtree$dev, type = "b") # lowest cv-error is when #nodes = 6
plot(cv.classtree$k, cv.classtree$dev, type = "b")
```

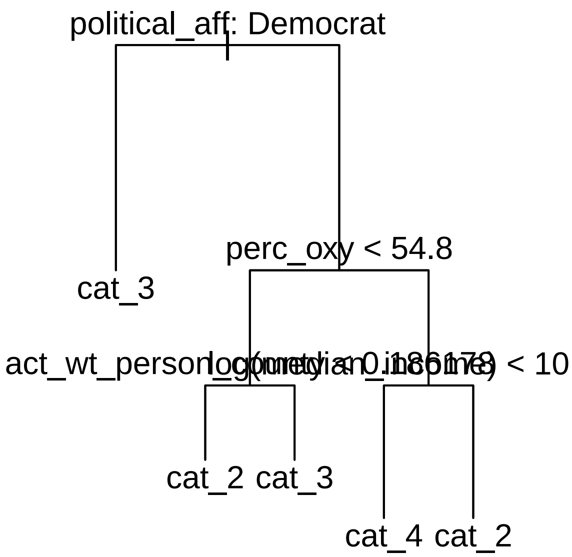



```
prune.classtree <- prune.misclass(classtree, best = 3)
plot(prune.classtree)
text(prune.classtree, pretty = 0)
prunetree.pred <- predict(prune.classtree, newdata = test_oh_wv_2012, type = "class")
table(prunetree.pred, deathrate.test)
```

```
##               deathrate.test
## prunetree.pred cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##          cat_1     0     0     0     0     0     0
##          cat_2     3    11     4     1     0     0
##          cat_3     0     5     3     0     2     5
##          cat_4     0     0     1     0     1     0
##          cat_5     0     0     0     0     0     0
##          cat_6     0     0     0     0     0     0
```

```r
sum(diag(table(prunetree.pred, deathrate.test)))/36 # correctly classified ~38.8%
```

```
## [1] 0.3888889
```




```r
## Bagging
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
set.seed(4)
bagtree <- randomForest(est_death_rate_cat ~., data = pred_matrix, mtry = 8, importance = TRUE, ntree =
bagtree
```

```
##
## Call:
##  randomForest(formula = est_death_rate_cat ~ ., data = pred_matrix,      mtry = 8, importance = TRUE
##                Type of random forest: classification
##                      Number of trees: 25
## No. of variables tried at each split: 8
##
##         OOB estimate of  error rate: 61.68%
## Confusion matrix:
##       cat_1 cat_2 cat_3 cat_4 cat_5 cat_6 class.error
## cat_1     0     6     1     0     0     0   1.0000000
## cat_2     3    26     8     4     2     0   0.3953488
## cat_3     0    14    11     3     1     1   0.6333333
```

```
## cat_4     0     4     8     4     1     1   0.7777778
## cat_5     0     1     2     2     0     0   1.0000000
## cat_6     0     0     0     3     1     0   1.0000000
```

```
bag.test <- test_oh_wv_2012 %>%
mutate(log_income = log(median_income)) %>%
dplyr::select(est_death_rate_cat, pharmacy_num_ptt, most_dist_channel, dominance, log_income, political_
bagtree.pred <- predict(bagtree, newdata = bag.test)
table(bagtree.pred, deathrate.test)
```

```
##             deathrate.test
## bagtree.pred cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##        cat_1     0     0     0     0     0     0
##        cat_2     3     9     4     1     0     0
##        cat_3     0     6     3     0     1     1
##        cat_4     0     0     1     0     1     3
##        cat_5     0     1     0     0     0     0
##        cat_6     0     0     0     0     1     1
```

```
sum(diag(table(bagtree.pred, deathrate.test)))/36 # correctly classified ~36%
```

```
## [1] 0.3611111
```

```
## RF
set.seed(5)
rf.tree <- randomForest(est_death_rate_cat ~., data = pred_matrix, mtry = 3, importance = TRUE)
rf.tree
```

```
##
## Call:
##  randomForest(formula = est_death_rate_cat ~ ., data = pred_matrix,      mtry = 3, importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 51.4%
## Confusion matrix:
##       cat_1 cat_2 cat_3 cat_4 cat_5 cat_6 class.error
## cat_1     0     7     0     0     0     0   1.0000000
## cat_2     1    33     6     3     0     0   0.2325581
## cat_3     0    12    13     3     2     0   0.5666667
## cat_4     0     7     5     6     0     0   0.6666667
## cat_5     0     1     4     0     0     0   1.0000000
## cat_6     0     0     2     2     0     0   1.0000000
```

```
rf.pred <- predict(rf.tree, newdata = bag.test)
table(rf.pred, deathrate.test)
```

```
##        deathrate.test
## rf.pred cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##   cat_1     0     0     0     0     0     0
##   cat_2     3    12     4     1     0     0
##   cat_3     0     3     2     0     1     1
##   cat_4     0     0     2     0     2     3
##   cat_5     0     1     0     0     0     0
##   cat_6     0     0     0     0     0     1
```

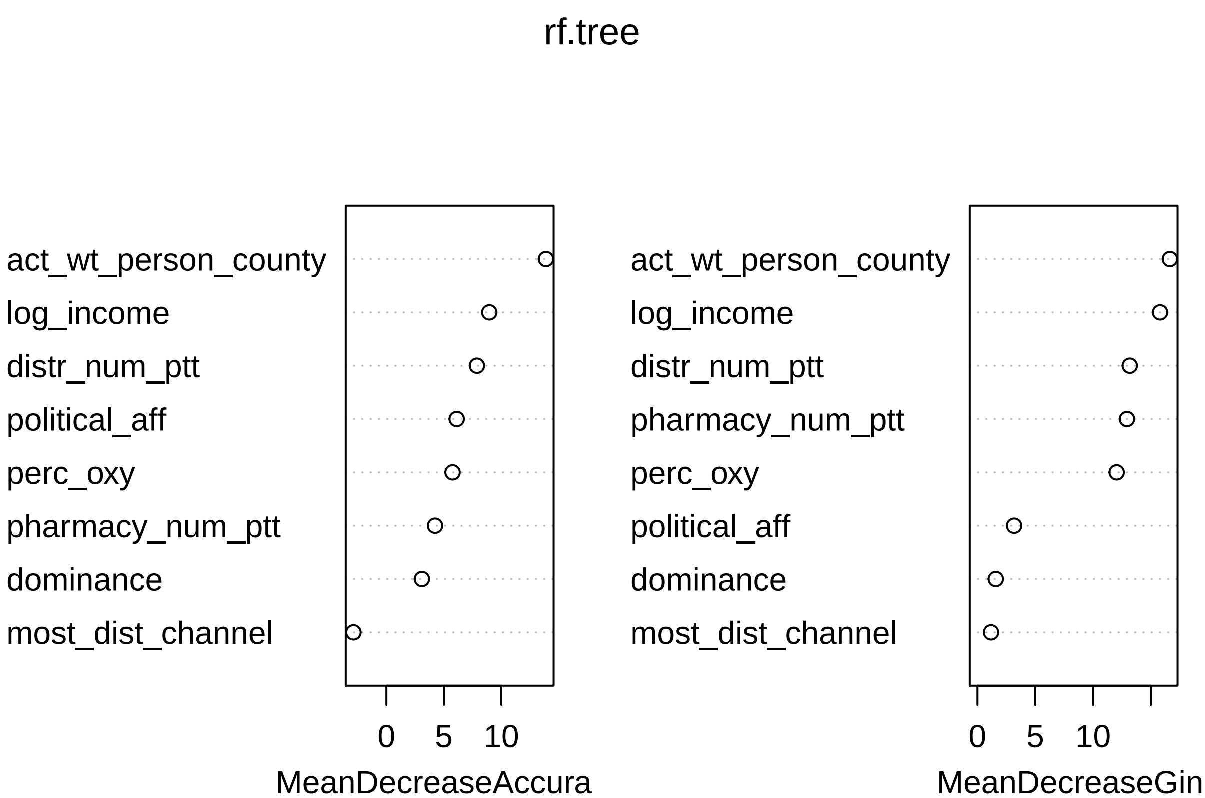
```r
sum(diag(table(rf.pred, deathrate.test)))/36 # correctly classified ~41.6%
```

```
## [1] 0.4166667
```

```r
importance(rf.tree) # log_income and act_wt_person are most important predictors
```

```
##                             cat_1     cat_2      cat_3      cat_4
## pharmacy_num_ptt       0.15538995  4.469156 -0.1031034  0.7077742
## most_dist_channel     -0.12840129 -4.374066  2.1196737 -1.0670856
## dominance              3.22947715  0.201426  1.5278826  4.0070936
## log_income             0.74744283 11.531918 -0.6112006  4.1277354
## political_aff          3.86518860  2.631297  4.8664872  0.1390327
## act_wt_person_county -0.22423245 10.881121  6.2096595  7.7694189
## perc_oxy              -0.03936824  6.192752  6.0994209 -2.2375869
## distr_num_ptt         -1.22333357  4.713947  6.1053007  3.1087917
##                               cat_5      cat_6 MeanDecreaseAccuracy
## pharmacy_num_ptt       4.680364e+00 -0.4487080             4.231039
## most_dist_channel      1.389425e+00 -1.3440623            -2.867095
## dominance              8.170415e-01 -0.3780185             3.084762
## log_income            -1.284557e+00 -2.9026957             8.951898
## political_aff          4.569845e+00  2.7593864             6.115234
## act_wt_person_county   2.162240e+00  1.9658005            13.878486
## perc_oxy              -3.348708e+00 -3.1288829             5.757402
## distr_num_ptt          5.496517e-17 -0.6030798             7.874999
##                      MeanDecreaseGini
## pharmacy_num_ptt            12.928056
## most_dist_channel            1.174849
## dominance                    1.583025
## log_income                  15.792193
## political_aff                3.168950
## act_wt_person_county        16.641734
## perc_oxy                    12.040750
## distr_num_ptt               13.173525
```

```r
varImpPlot(rf.tree)
```



Ordinal package

```
library(ordinalNet)
y<-as.factor(train_oh_wv_2012$est_death_rate_cat)
x<-model.matrix(est_death_rate_cat~., pred_matrix)
#View(x)
ordnet1 <- ordinalNet(x, y, family="cumulative",
                    parallelTerms=FALSE, nonparallelTerms=TRUE, # alpha = 1 means Lasso
                    standardize = FALSE)
```

```
## Warning in ordinalNet(x, y, family = "cumulative", parallelTerms =
## FALSE, : For out-of-sample data, the cumulative probability model with
## nonparallelTerms=TRUE may predict cumulative probabilities that are not
## monotone increasing.
```

```
summary(ordnet1)
```

```
##    lambdaVals nNonzero    loglik      devPct      aic      bic
## 1  1.55122718        5 -156.9847 0.000000000 323.9695 337.3336
## 2  1.21734100        7 -155.9926 0.006319662 325.9853 344.6951
## 3  0.95532048        7 -154.4556 0.016110402 322.9113 341.6211
## 4  0.74969728        7 -153.6019 0.021548960 321.2037 339.9135
## 5  0.58833241        7 -153.5762 0.021712192 321.1525 339.8623
## 6  0.46169973        7 -153.5762 0.021712247 321.1525 339.8623
## 7  0.36232346        7 -153.5762 0.021712253 321.1525 339.8623
## 8  0.28433695        7 -153.5762 0.021712253 321.1525 339.8623
## 9  0.22313626        7 -153.5762 0.021712253 321.1525 339.8623
## 10 0.17510840        7 -153.5762 0.021712253 321.1525 339.8623
```

```
## 11 0.13741806        7 -153.5762 0.021712253 321.1525 339.8623
## 12 0.10784020        7 -153.5762 0.021712253 321.1525 339.8623
## 13 0.08462867        7 -153.5762 0.021712253 321.1525 339.8623
## 14 0.06641319        7 -153.5762 0.021712253 321.1525 339.8623
## 15 0.05211841        7 -153.5762 0.021712253 321.1525 339.8623
## 16 0.04090045        7 -153.5762 0.021712253 321.1525 339.8623
## 17 0.03209703        7 -153.5762 0.021712253 321.1525 339.8623
## 18 0.02518847        7 -153.5762 0.021712253 321.1525 339.8623
## 19 0.01976690        7 -153.5762 0.021712253 321.1525 339.8623
## 20 0.01551227        7 -153.5762 0.021712253 321.1525 339.8623
```

```r
coef(ordnet1, matrix=TRUE, criteria="aic") #by default, best AIC model is returned
```

```
##                                  logit(P[Y<=1]) logit(P[Y<=2])
## (Intercept)                           -2.656564     -0.1262947
## (Intercept)                            0.000000      0.0000000
## pharmacy_num_ptt                       0.000000      0.0000000
## most_dist_channelRETAIL PHARMACY       0.000000      0.0000000
## dominanceYes                           0.000000      0.0000000
## log_income                             0.000000      0.0000000
## political_affRepublican                0.000000      0.0000000
## act_wt_person_county                   0.000000      0.0000000
## perc_oxy                               0.000000      0.0000000
## distr_num_ptt                          0.000000      0.0000000
##                                  logit(P[Y<=3]) logit(P[Y<=4])
## (Intercept)                         1.625674484     1.11795827
## (Intercept)                         0.000000000     0.00000000
## pharmacy_num_ptt                    0.000000000     0.00000000
## most_dist_channelRETAIL PHARMACY    0.000000000     0.00000000
## dominanceYes                        0.000000000     0.00000000
## log_income                          0.000000000     0.00000000
## political_affRepublican             0.000000000     0.00000000
## act_wt_person_county                0.000000000     0.00000000
## perc_oxy                           -0.008977092     0.02071391
## distr_num_ptt                       0.000000000     0.00000000
##                                  logit(P[Y<=5])
## (Intercept)                            3.050339
## (Intercept)                            0.000000
## pharmacy_num_ptt                       0.000000
## most_dist_channelRETAIL PHARMACY       0.000000
## dominanceYes                           0.000000
## log_income                             0.000000
## political_affRepublican                0.000000
## act_wt_person_county                   0.000000
## perc_oxy                               0.000000
## distr_num_ptt                          0.000000
```

```r
# CV by misclassification error
# ordinalNetCV(x, y, tuneMethod = "cvMisclass")
```

Calculate accuracy for all logistic regression models

fit1_interact_ord

```r
deathrate.test <- test_oh_wv_2012$est_death_rate_cat
classtree.pred.fit1_interact_ord <- predict(fit1_interact_ord, test_oh_wv_2012, type = "class")
table(classtree.pred.fit1_interact_ord, deathrate.test)
```

```
##                                 deathrate.test
## classtree.pred.fit1_interact_ord cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##                            cat_1     0     3     0     0     0     0
##                            cat_2     3    10     3     0     0     0
##                            cat_3     0     0     2     1     1     0
##                            cat_4     0     0     2     0     1     1
##                            cat_5     0     2     1     0     0     0
##                            cat_6     0     1     0     0     1     4
```

```r
sum(diag(table(classtree.pred.fit1_interact_ord, deathrate.test)))/36
```

```
## [1] 0.4444444
```

fit.select

```r
deathrate.test <- test_oh_wv_2012$est_death_rate_cat
fit.select.preds <- predict(fit.select, test_oh_wv_2012, type = "class")
table(fit.select.preds, deathrate.test)
```

```
##                 deathrate.test
## fit.select.preds cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##            cat_1     1     0     0     0     0     0
##            cat_2     2    13     3     0     0     0
##            cat_3     0     2     5     1     2     0
##            cat_4     0     1     0     0     1     2
##            cat_5     0     0     0     0     0     0
##            cat_6     0     0     0     0     0     3
```

```r
sum(diag(table(fit.select.preds, deathrate.test)))/36
```

```
## [1] 0.6111111
```

fit0_interact

```r
deathrate.test <- test_oh_wv_2012$est_death_rate_cat
classtree.pred.fit0.interact <- predict(fit0.interact, test_oh_wv_2012, type = "class")
table(classtree.pred.fit0.interact, deathrate.test)
```

```
##                             deathrate.test
## classtree.pred.fit0.interact cat_1 cat_2 cat_3 cat_4 cat_5 cat_6
##                        cat_1     0     3     0     0     0     0
##                        cat_2     3    10     3     0     0     0
##                        cat_3     0     0     2     1     1     0
##                        cat_4     0     0     2     0     1     1
##                        cat_5     0     2     1     0     0     0
##                        cat_6     0     1     0     0     1     4
```

```r
sum(diag(table(classtree.pred.fit0.interact, deathrate.test)))/36
```

```
## [1] 0.4444444
```

ordnet

```
pred_matrix.ordnet1 <- test_oh_wv_2012 %>%  #x is a matrix of predictors
  dplyr::select(est_death_rate_cat, pharmacy_num_ptt, most_dist_channel, dominance, median_income,
                political_aff, act_wt_person_county, perc_oxy, distr_num_ptt)
x.test.ordnet1<-model.matrix(est_death_rate_cat~., pred_matrix.ordnet1)
ordnet.pred<-predict(ordnet1, newx = x.test.ordnet1, whichLambda = NULL,
      criteria = "aic", type = "class")
deathrate.test <- as.character(test_oh_wv_2012$est_death_rate_cat)
no.cat.test<-str_replace(deathrate.test, "cat_", "")
no.cat.test<-as.numeric(no.cat.test)
table(ordnet.pred, no.cat.test)
```

```
##            no.cat.test
## ordnet.pred  1  2  3  4  5  6
##           2  3 16  8  1  3  5
```

```
z<-c(0, 0, 0, 0, 0)
table<-rbind(table(ordnet.pred, no.cat.test), z)
```

```
## Warning in rbind(table(ordnet.pred, no.cat.test), z): number of columns of
## result is not a multiple of vector length (arg 2)
```

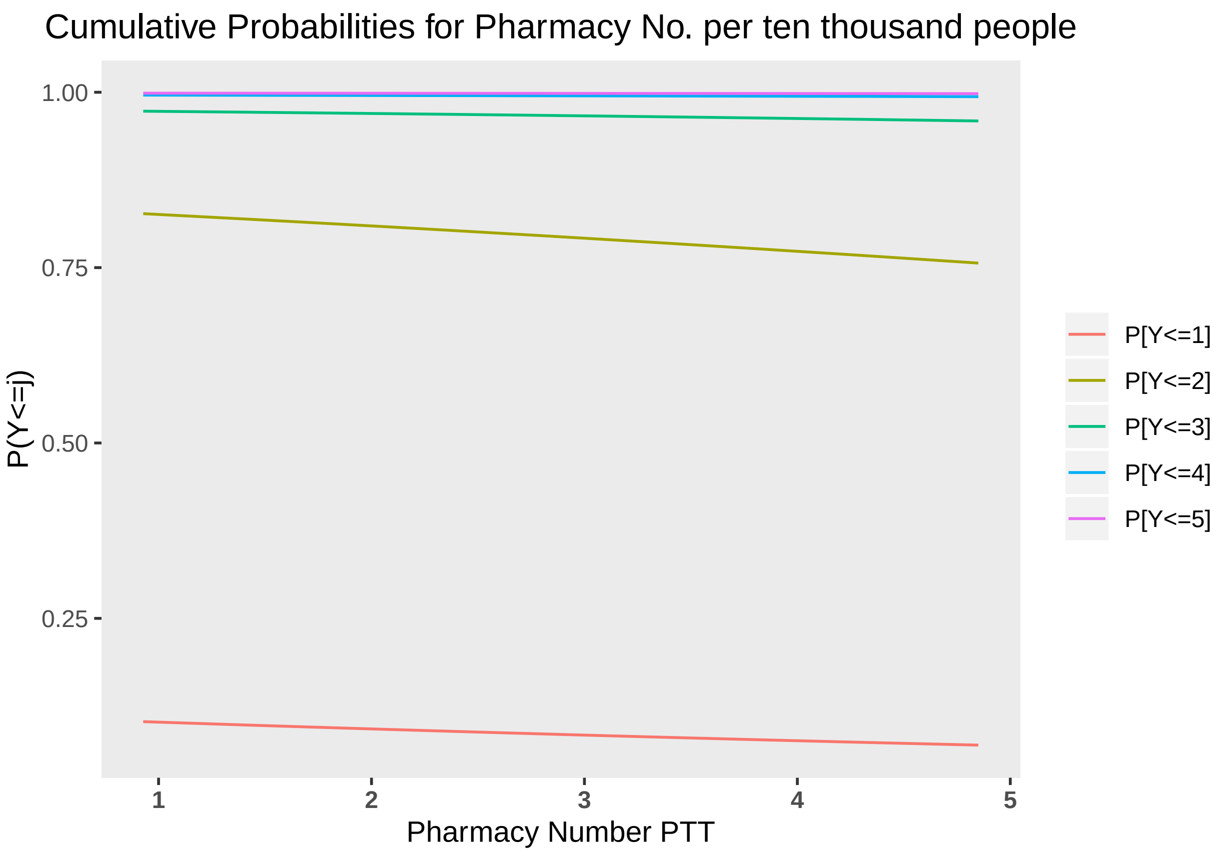
```
sum(diag(table))/36
```

```
## [1] 0.08333333
```

prediction plots for fit.select using TRAINING

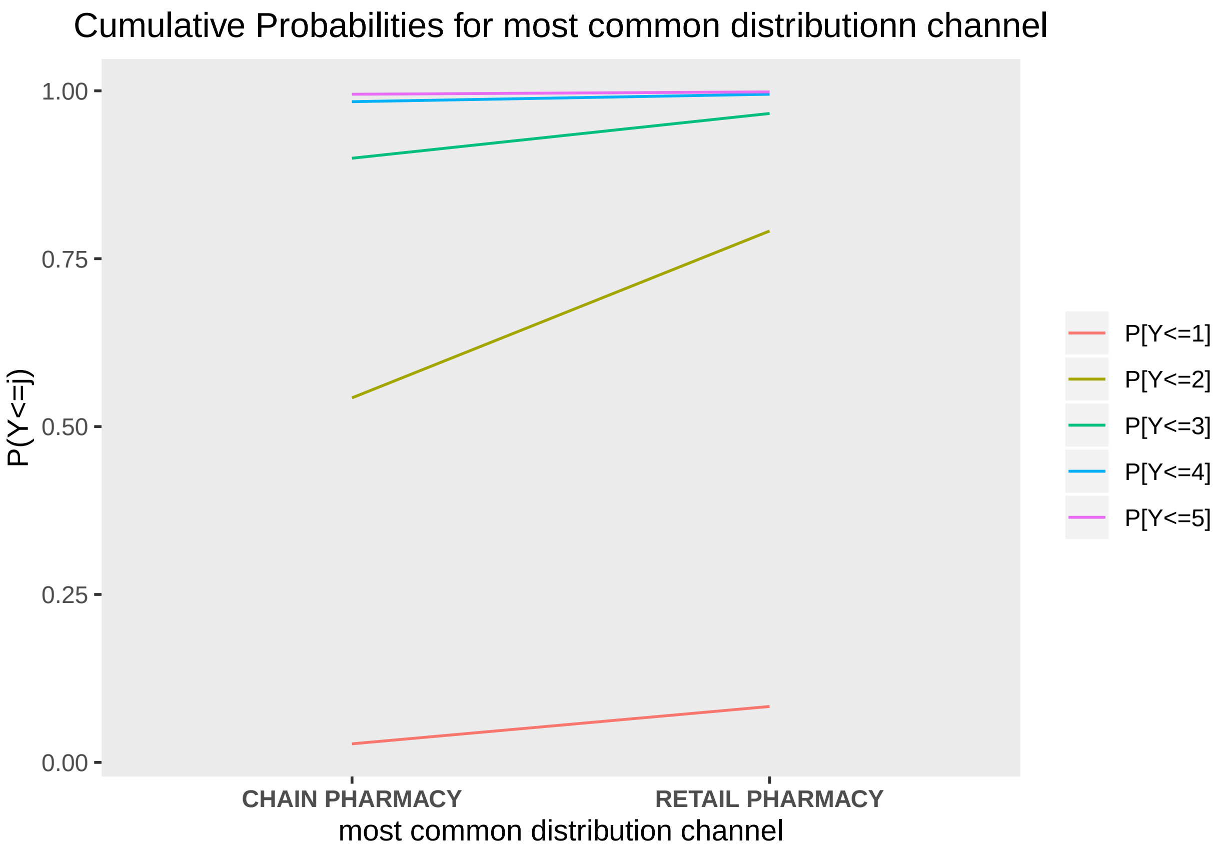
```
#pharmacy no
pharm_num.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_pharm_num  <- predict(fit.select, newdata = pharm_num.test.ordnet1, type = "probs")
# plotting
classprob_pharm_num_df = t(classprob_pharm_num) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_pharm_num_df) = NULL
colnames(classprob_pharm_num_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_pharm_num_df = as.data.frame(classprob_pharm_num_df) %>%
  cbind(pharm_num.test.ordnet1) %>%
  dplyr::select(pharmacy_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_pharm_num_df, aes(x = pharmacy_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Pharmacy No. per ten thousand people",
         y = "P(Y<=j)",
         x= "Pharmacy Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
```

```r
        axis.text.x = element_text(face = "bold"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank())
```



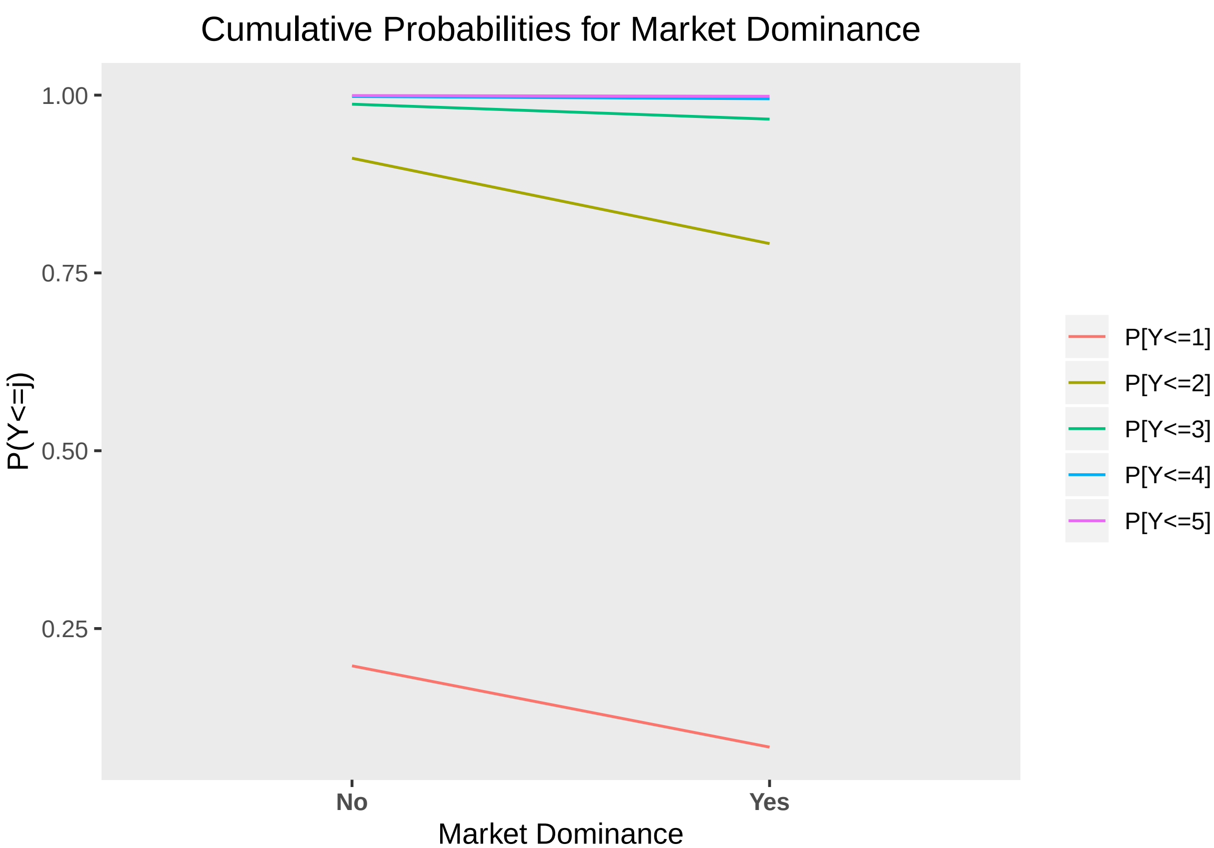

```r
#summary(fit.select)
# most_dist_channel
dom_channel.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, dominance="Yes", median_income=43194.13,
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_most_dist_channel <- predict(fit.select, newdata = dom_channel.test.ordnet1, type = "probs")
# plotting
classprob_most_dist_channel_df = t(classprob_most_dist_channel) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_most_dist_channel_df) = NULL
colnames(classprob_most_dist_channel_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_most_dist_channel_df = as.data.frame(classprob_most_dist_channel_df) %>%
  cbind(dom_channel.test.ordnet1) %>%
  dplyr::select(most_dist_channel, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         most_dist_channel = as.factor(most_dist_channel))
```

```r
ggplot(classcumprob_most_dist_channel_df, aes(x = most_dist_channel, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for most common distributionn channel",
         y = "P(Y<=j)",
         x= "most common distribution channel") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



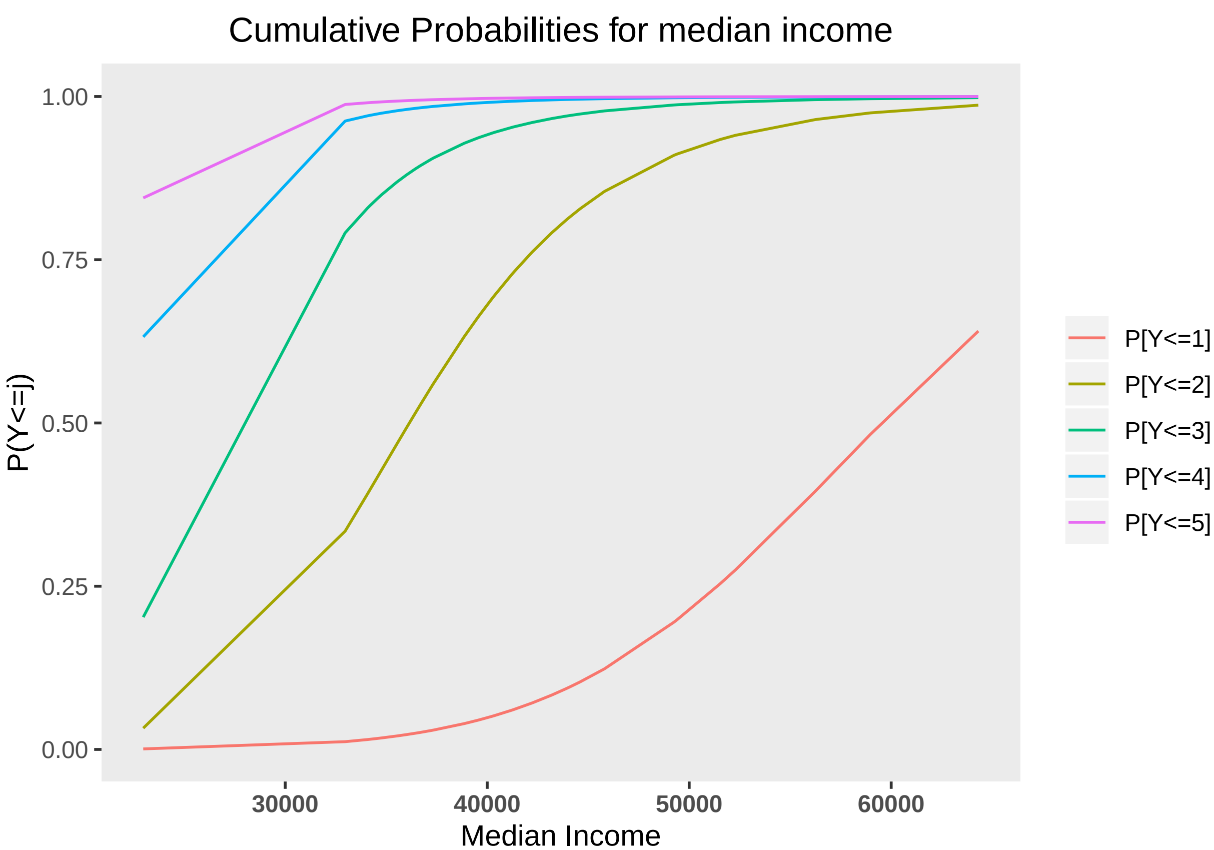

```r
# dominance
dom_channel.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", median_income=43194.13,
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_dom <- predict(fit.select, newdata = dom_channel.test.ordnet1, type = "probs")
# plotting
classprob_dom_df = t(classprob_dom) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_dom_df) = NULL
colnames(classprob_dom_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
```

```
classcumprob_dom_df = as.data.frame(classprob_dom_df) %>%
  cbind(dom_channel.test.ordnet1) %>%
  dplyr::select(dominance, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         dominance = as.factor(dominance))

ggplot(classcumprob_dom_df, aes(x = dominance, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Market Dominance",
         y = "P(Y<=j)",
         x= "Market Dominance") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



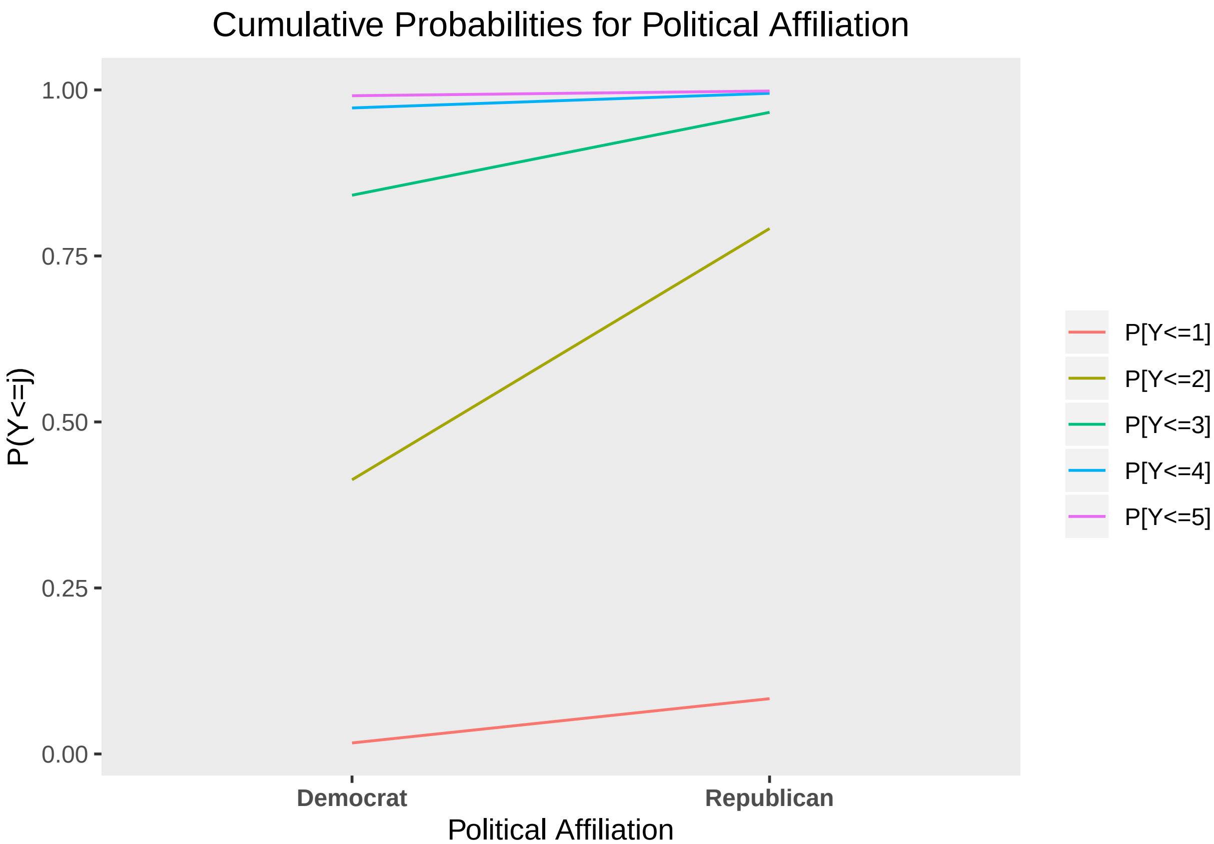

```
# income
log_income.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_log_income <- predict(fit.select, newdata = log_income.test.ordnet1, type = "probs")
# plotting
classprob_log_income_df = t(classprob_log_income) %>%
```

```
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_log_income_df) = NULL
colnames(classprob_log_income_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_log_income_df = as.data.frame(classprob_log_income_df) %>%
  cbind(log_income.test.ordnet1) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_log_income_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for median income",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



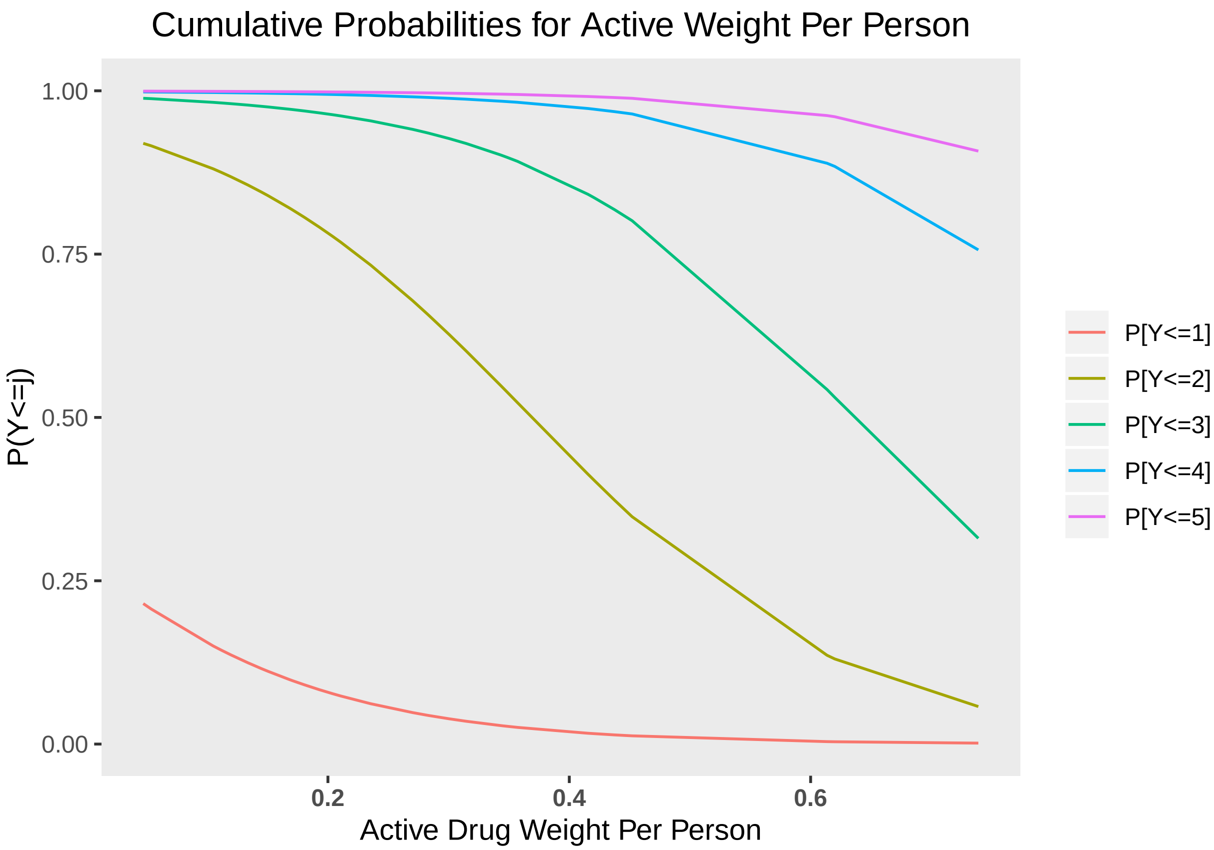

```
# political_aff
polaff.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
    mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
```

```r
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_polaff <- predict(fit.select, newdata = polaff.test.ordnet1, type = "probs")
# plotting
classprob_polaff_df = t(classprob_polaff) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_polaff_df) = NULL
colnames(classprob_polaff_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_polaff_df = as.data.frame(classprob_polaff_df) %>%
  cbind(polaff.test.ordnet1) %>%
  dplyr::select(political_aff, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         political_aff = as.factor(political_aff))

ggplot(classcumprob_polaff_df, aes(x = political_aff, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Political Affiliation",
         y = "P(Y<=j)",
         x= "Political Affiliation") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



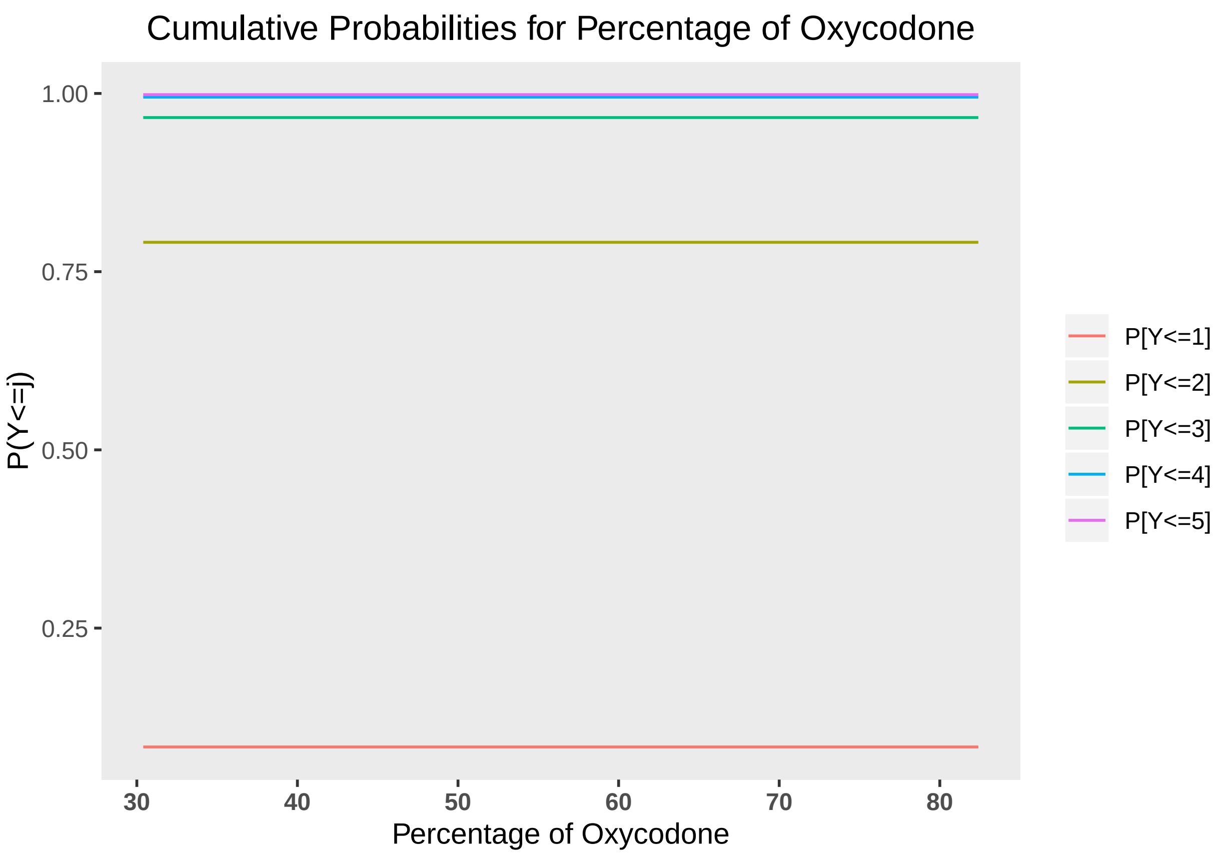

```r
# act_wt_person_county
act_wt.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_act_wt <- predict(fit.select, newdata = act_wt.test.ordnet1, type = "probs")
# plotting
classprob_act_wt_df = t(classprob_act_wt) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_act_wt_df) = NULL
colnames(classprob_act_wt_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_act_wt_df = as.data.frame(classprob_act_wt_df) %>%
  cbind(act_wt.test.ordnet1) %>%
  dplyr::select(act_wt_person_county, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_act_wt_df, aes(x = act_wt_person_county, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Active Weight Per Person",
         y = "P(Y<=j)",
         x= "Active Drug Weight Per Person") +
    theme(plot.title = element_text(hjust = 0.5),
```

```r
        axis.text.x = element_text(face = "bold"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank())
```



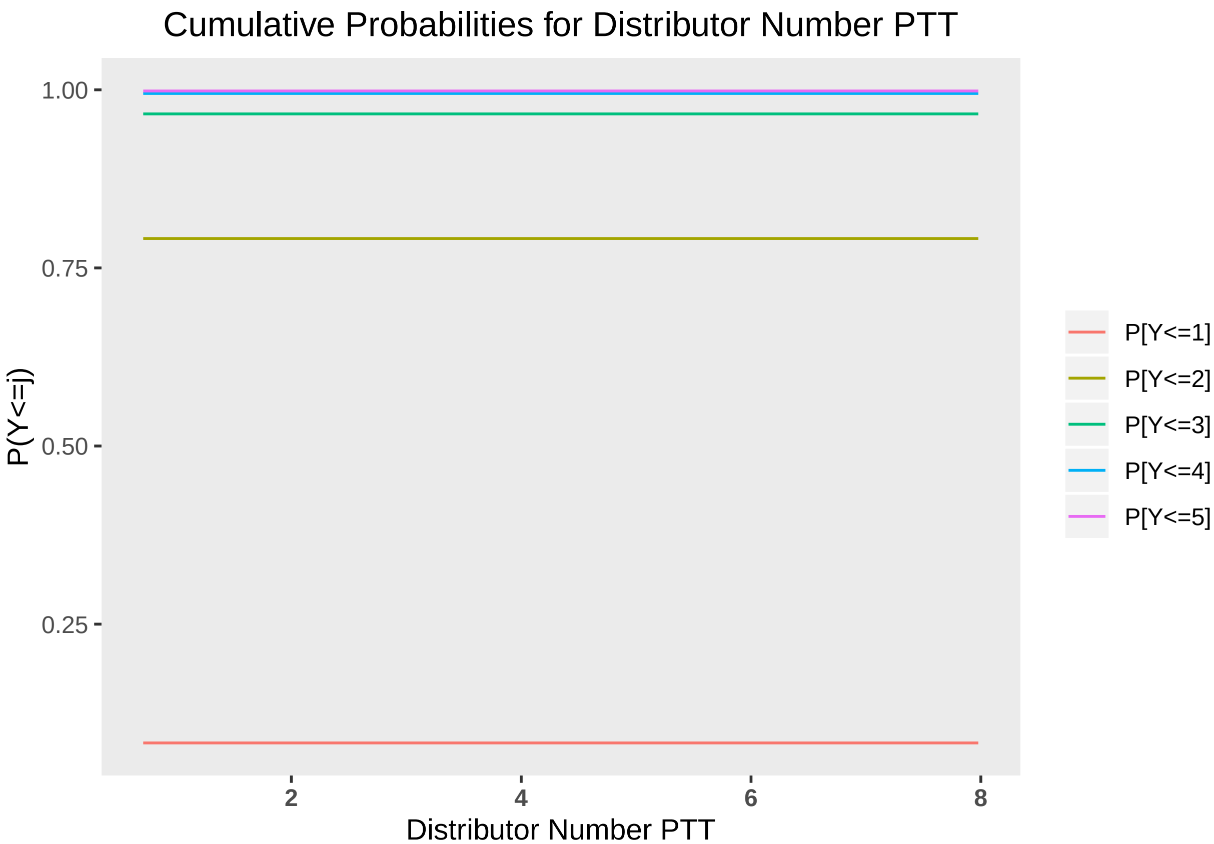

```r
# perc_oxy --> NOT SIGNIFICANT IN CURRENT MODEL
perc_oxy.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_perc_oxy <- predict(fit.select, newdata = perc_oxy.test.ordnet1, type = "probs")
# plotting
classprob_perc_oxy_df = t(classprob_perc_oxy) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_perc_oxy_df) = NULL
colnames(classprob_perc_oxy_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_perc_oxy_df = as.data.frame(classprob_perc_oxy_df) %>%
  cbind(perc_oxy.test.ordnet1) %>%
  dplyr::select(perc_oxy, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_perc_oxy_df, aes(x = perc_oxy, y = probability)) +
```

```
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Percentage of Oxycodone",
         y = "P(Y<=j)",
         x= "Percentage of Oxycodone") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```




```
# distr_num_ptt
distr_num.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         act_wt_person_county = 0.19294084)
classprob_distr_num  <- predict(fit.select, newdata = distr_num.test.ordnet1, type = "probs")
# plotting
classprob_distr_num_df = t(classprob_distr_num) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_distr_num_df) = NULL
colnames(classprob_distr_num_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_distr_num_df = as.data.frame(classprob_distr_num_df) %>%
  cbind(distr_num.test.ordnet1) %>%
```

```
  dplyr::select(distr_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_distr_num_df, aes(x = distr_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Distributor Number PTT",
         y = "P(Y<=j)",
         x= "Distributor Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```

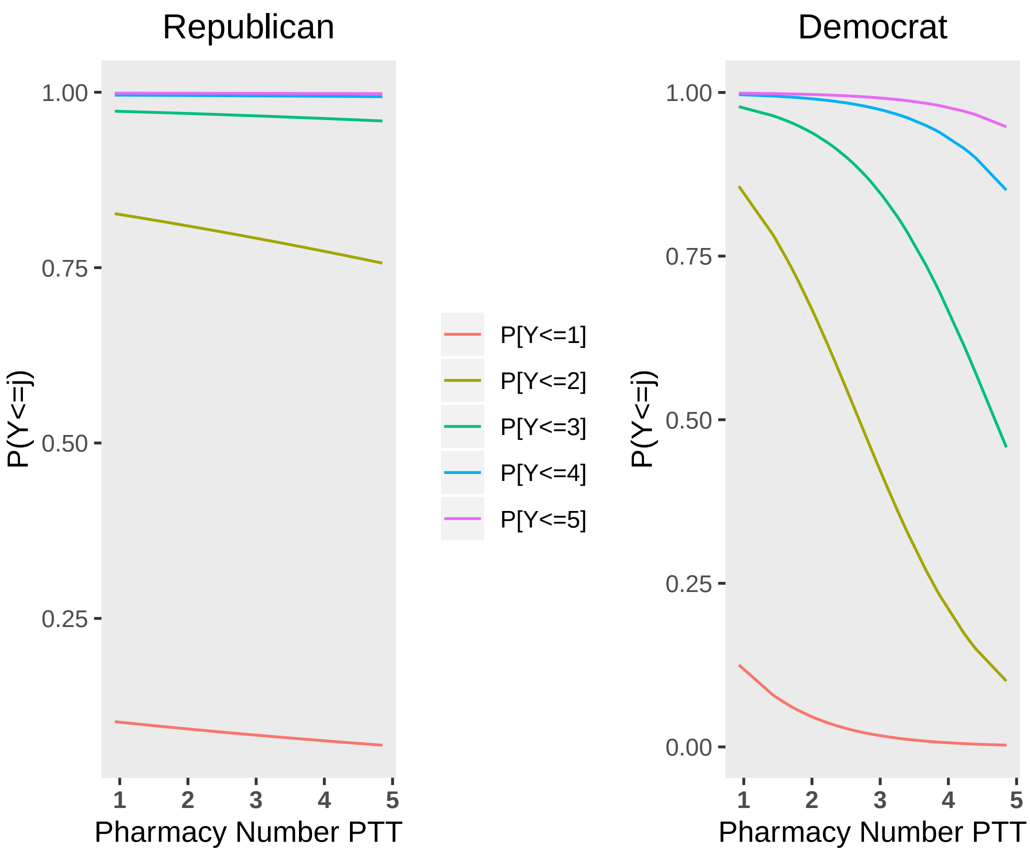


interaction plots for fit.select ON TESTING

```
## number of pharmacies and political affilition
# Republican
pharm_num.test.ordnet1_rep = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_pharm_num_rep  <- predict(fit.select, newdata = pharm_num.test.ordnet1_rep, type = "probs", s
# plotting
```

```r
classprob_pharm_num_rep_df = t(classprob_pharm_num_rep) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_pharm_num_rep_df) = NULL
colnames(classprob_pharm_num_rep_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_pharm_num_rep_df = as.data.frame(classprob_pharm_num_rep_df) %>%
  cbind(pharm_num.test.ordnet1_rep) %>%
  dplyr::select(pharmacy_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

num_pharm_rep_p = ggplot(classcumprob_pharm_num_rep_df, aes(x = pharmacy_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Republican",
         y = "P(Y<=j)",
         x= "Pharmacy Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
# Democrat
pharm_num.test.ordnet1_dem = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Democrat",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_pharm_num_dem  <- predict(fit.select, newdata = pharm_num.test.ordnet1_dem, type = "probs", s
# plotting
classprob_pharm_num_dem_df = t(classprob_pharm_num_dem) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_pharm_num_dem_df) = NULL
colnames(classprob_pharm_num_dem_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_pharm_num_dem_df = as.data.frame(classprob_pharm_num_dem_df) %>%
  cbind(pharm_num.test.ordnet1_dem) %>%
  dplyr::select(pharmacy_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

num_pharm_dem_p = ggplot(classcumprob_pharm_num_dem_df, aes(x = pharmacy_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Democrat",
         y = "P(Y<=j)",
         x= "Pharmacy Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
```

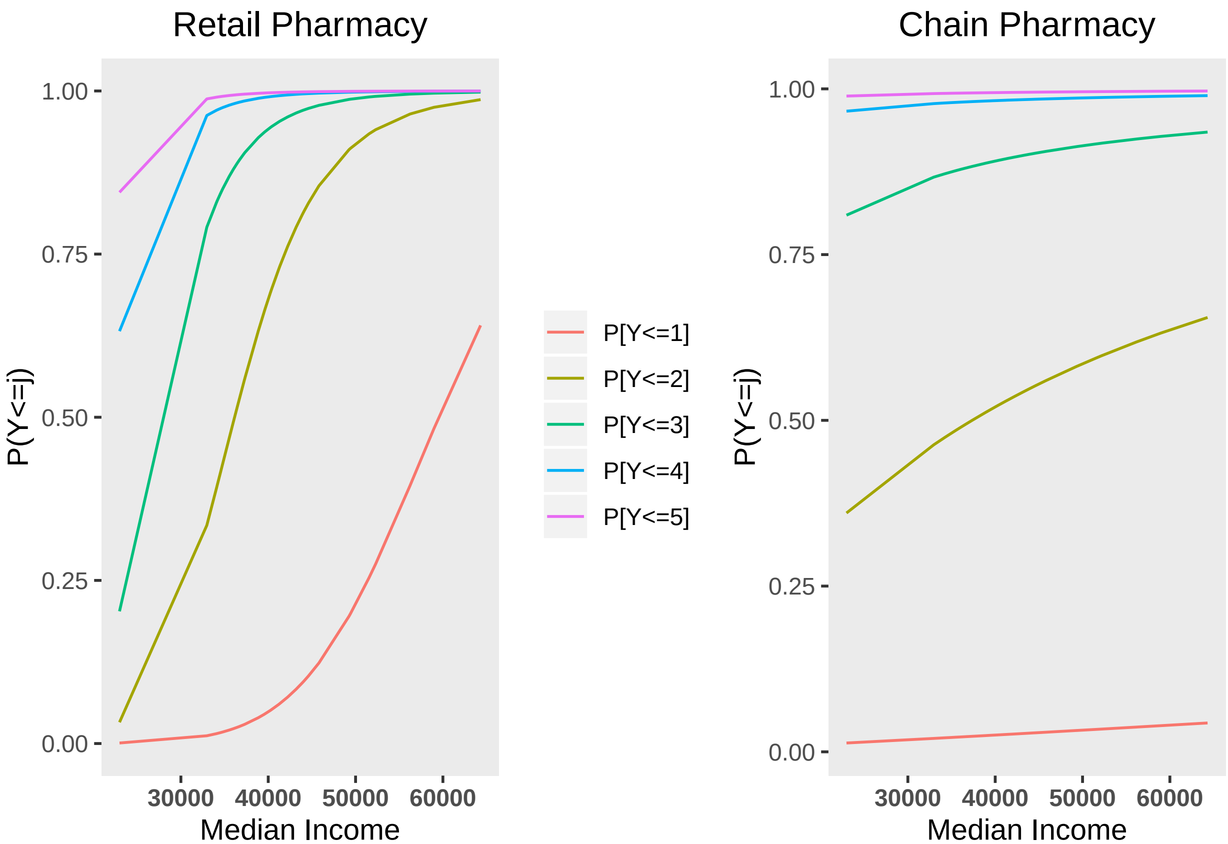
```
          legend.title = element_blank(),
          legend.position = "none")
# num_pharm_rep_p + num_pharm_dem_p
plot_grid(num_pharm_rep_p, num_pharm_dem_p, axis = "r", align = "v")
```




```
# probability of falling below a category decreases faster for democrats. is
# significant? because our dataset is small.
```

```
## distribution channel and income
# RETAIL PHARMACRY
income.test.ordnet1_rp = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 3.038,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_rp  <- predict(fit.select, newdata = income.test.ordnet1_rp, type = "probs", se.fit = '
# plotting
classprob_income_rp_df = t(classprob_income_rp) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_rp_df) = NULL
colnames(classprob_income_rp_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_rp_df = as.data.frame(classprob_income_rp_df) %>%
  cbind(income.test.ordnet1_rp) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
```

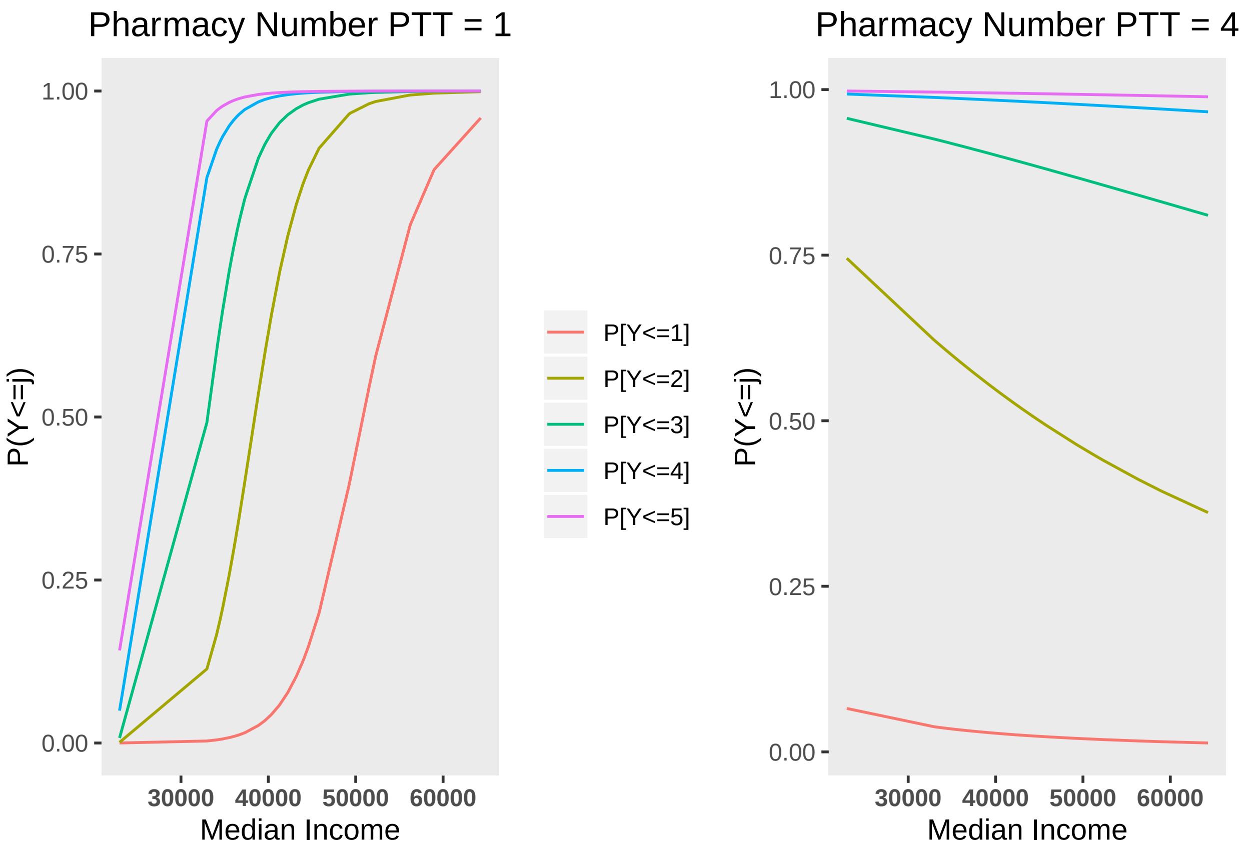
```r
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_rp_p = ggplot(classcumprob_income_rp_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Retail Pharmacy",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
# RETAIL PHARMACRY
income.test.ordnet1_cp = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "CHAIN PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 3.038,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_cp  <- predict(fit.select, newdata = income.test.ordnet1_cp, type = "probs", se.fit = 
# plotting
classprob_income_cp_df = t(classprob_income_cp) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_cp_df) = NULL
colnames(classprob_income_cp_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_cp_df = as.data.frame(classprob_income_cp_df) %>%
  cbind(income.test.ordnet1_cp) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_cp_p = ggplot(classcumprob_income_cp_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Chain Pharmacy",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          legend.position = "none")
income_rp_p + income_cp_p
```




```
# plot_grid(income_rp_p, income_cp_p, axis = "r", align = "v")
```

```
## pharmacy number and log median income
# range of pharmacy number ptt: 0.928 4.850
# fixing pharmacy number at = 1
income.test.ordnet1_1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 1,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_1  <- predict(fit.select, newdata = income.test.ordnet1_1, type = "probs", se.fit = TRU
# plotting
classprob_income_1_df = t(classprob_income_1) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_1_df) = NULL
colnames(classprob_income_1_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_1_df = as.data.frame(classprob_income_1_df) %>%
  cbind(income.test.ordnet1_1) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_1_p = ggplot(classcumprob_income_1_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
```

```
    labs(title = "Pharmacy Number PTT = 1",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
# fixing pharmacy number at = 4
income.test.ordnet1_4 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "CHAIN PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 4,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_4  <- predict(fit.select, newdata = income.test.ordnet1_4, type = "probs", se.fit = TRU
# plotting
classprob_income_4_df = t(classprob_income_4) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_4_df) = NULL
colnames(classprob_income_4_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_4_df = as.data.frame(classprob_income_4_df) %>%
  cbind(income.test.ordnet1_4) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_4_p = ggplot(classcumprob_income_4_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Pharmacy Number PTT = 4",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          legend.position = "none")
income_1_p + income_4_p
```

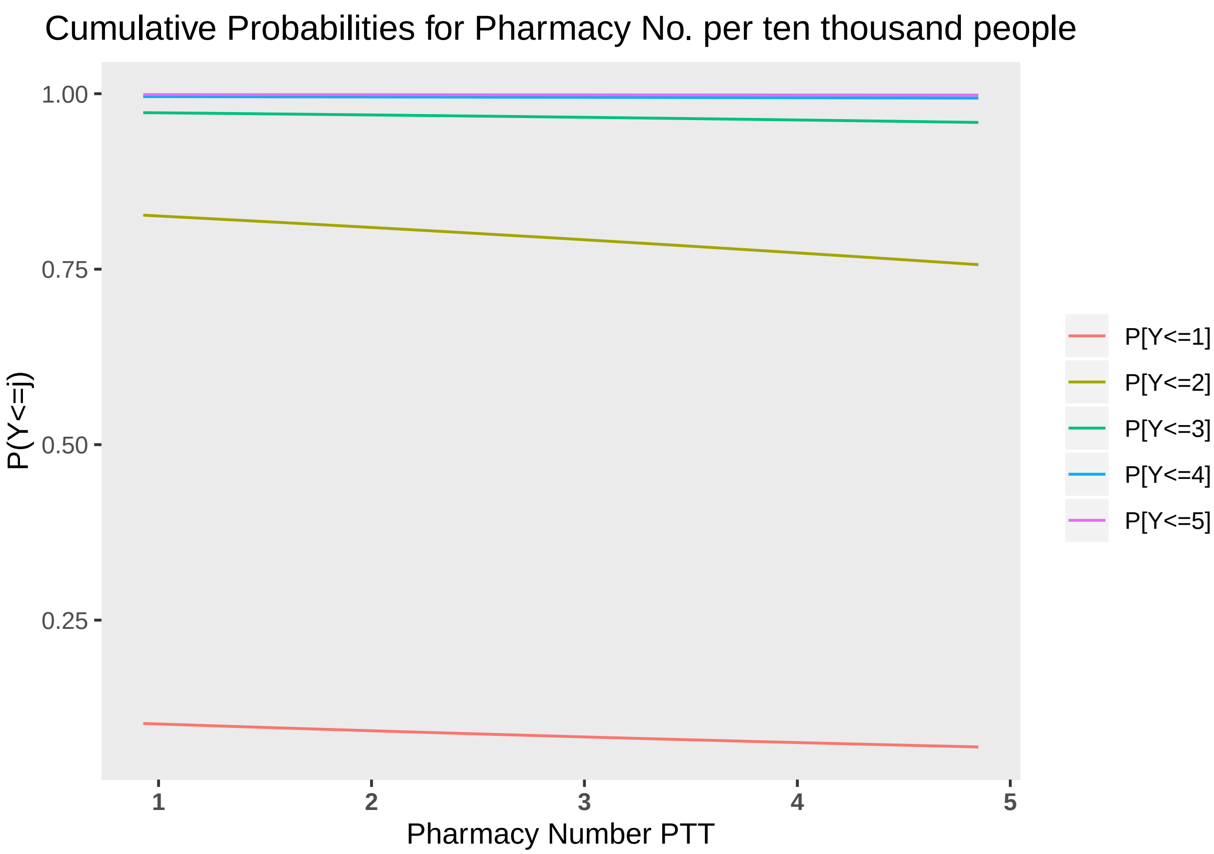



```
# plot_grid(income_1_p, income_4_p, axis = "r", align = "v")
```

prediction plots for fit.select ON TESTING

```
#pharmacy no
pharm_num.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_pharm_num  <- predict(fit.select, newdata = pharm_num.test.ordnet1, type = "probs")
# plotting
classprob_pharm_num_df = t(classprob_pharm_num) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_pharm_num_df) = NULL
colnames(classprob_pharm_num_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_pharm_num_df = as.data.frame(classprob_pharm_num_df) %>%
  cbind(pharm_num.test.ordnet1) %>%
  dplyr::select(pharmacy_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_pharm_num_df, aes(x = pharmacy_num_ptt, y = probability)) +
```

```r
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Pharmacy No. per ten thousand people",
         y = "P(Y<=j)",
         x= "Pharmacy Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



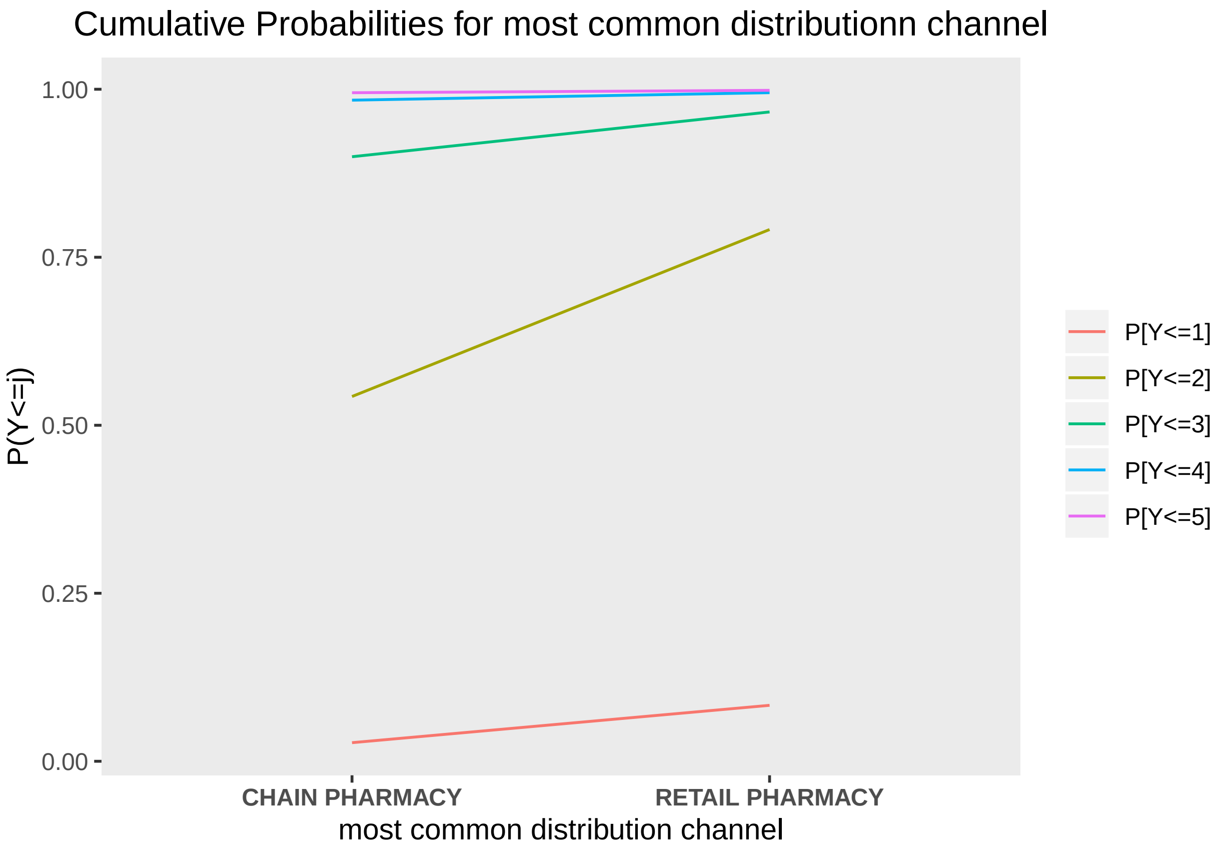

```r
summary(fit.select)
```

```
## Call:
## polr(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     pharmacy_num_ptt:political_aff + most_dist_channel:log(median_income) +
##     pharmacy_num_ptt:log(median_income), data = train_oh_wv_2012,
##     Hess = TRUE, method = "logistic")
##
## Coefficients:
##                                                       Value Std. Error
## pharmacy_num_ptt                                   -29.8071    13.1492
## most_dist_channelRETAIL PHARMACY                    66.0704    27.7312
## dominanceYes                                         0.9971     0.5648
## log(median_income)                                  -9.9542     3.4002
## political_affRepublican                              1.0636     1.3325
## act_wt_person_county                                 7.5644     1.7395
```

```
## pharmacy_num_ptt:political_affRepublican            -0.9046     0.4408
## most_dist_channelRETAIL PHARMACY:log(median_income)  -6.2989     2.6236
## pharmacy_num_ptt:log(median_income)                   2.8877     1.2485
##                                                     t value
## pharmacy_num_ptt                                    -2.2668
## most_dist_channelRETAIL PHARMACY                     2.3825
## dominanceYes                                         1.7653
## log(median_income)                                  -2.9275
## political_affRepublican                              0.7982
## act_wt_person_county                                 4.3485
## pharmacy_num_ptt:political_affRepublican            -2.0523
## most_dist_channelRETAIL PHARMACY:log(median_income) -2.4009
## pharmacy_num_ptt:log(median_income)                  2.3129
##
## Intercepts:
##             Value     Std. Error t value
## cat_1|cat_2 -105.9514   36.0762    -2.9369
## cat_2|cat_3 -102.2201   35.9160    -2.8461
## cat_3|cat_4 -100.1991   35.8594    -2.7942
## cat_4|cat_5  -98.2888   35.8476    -2.7419
## cat_5|cat_6  -97.1360   35.8385    -2.7104
##
## Residual Deviance: 236.3985
## AIC: 264.3985
```

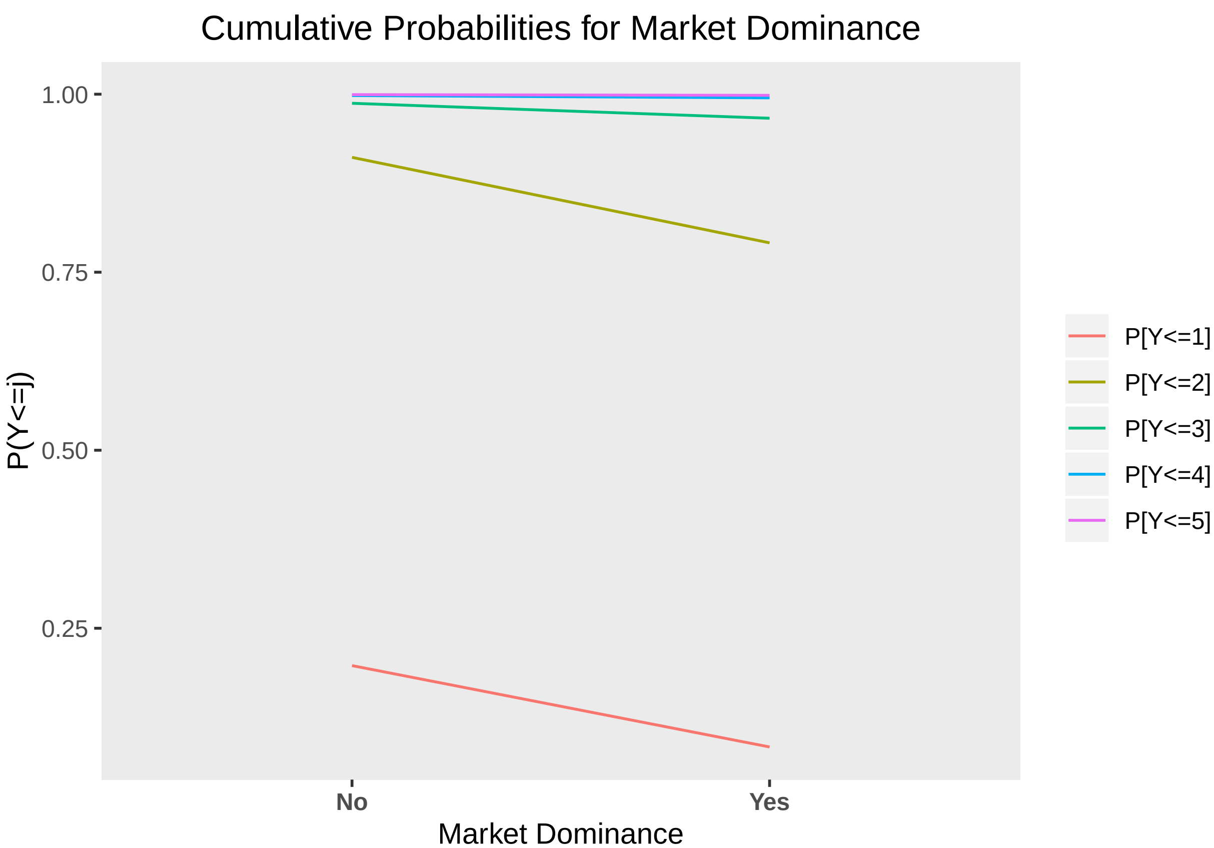
```r
# most_dist_channel
dom_channel.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, dominance="Yes", median_income=43194.13,
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_most_dist_channel <- predict(fit.select, newdata = dom_channel.test.ordnet1, type = "probs")
# plotting
classprob_most_dist_channel_df = t(classprob_most_dist_channel) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_most_dist_channel_df) = NULL
colnames(classprob_most_dist_channel_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_most_dist_channel_df = as.data.frame(classprob_most_dist_channel_df) %>%
  cbind(dom_channel.test.ordnet1) %>%
  dplyr::select(most_dist_channel, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         most_dist_channel = as.factor(most_dist_channel))

ggplot(classcumprob_most_dist_channel_df, aes(x = most_dist_channel, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for most common distributionn channel",
         y = "P(Y<=j)",
         x= "most common distribution channel") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
```

```
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank())
```



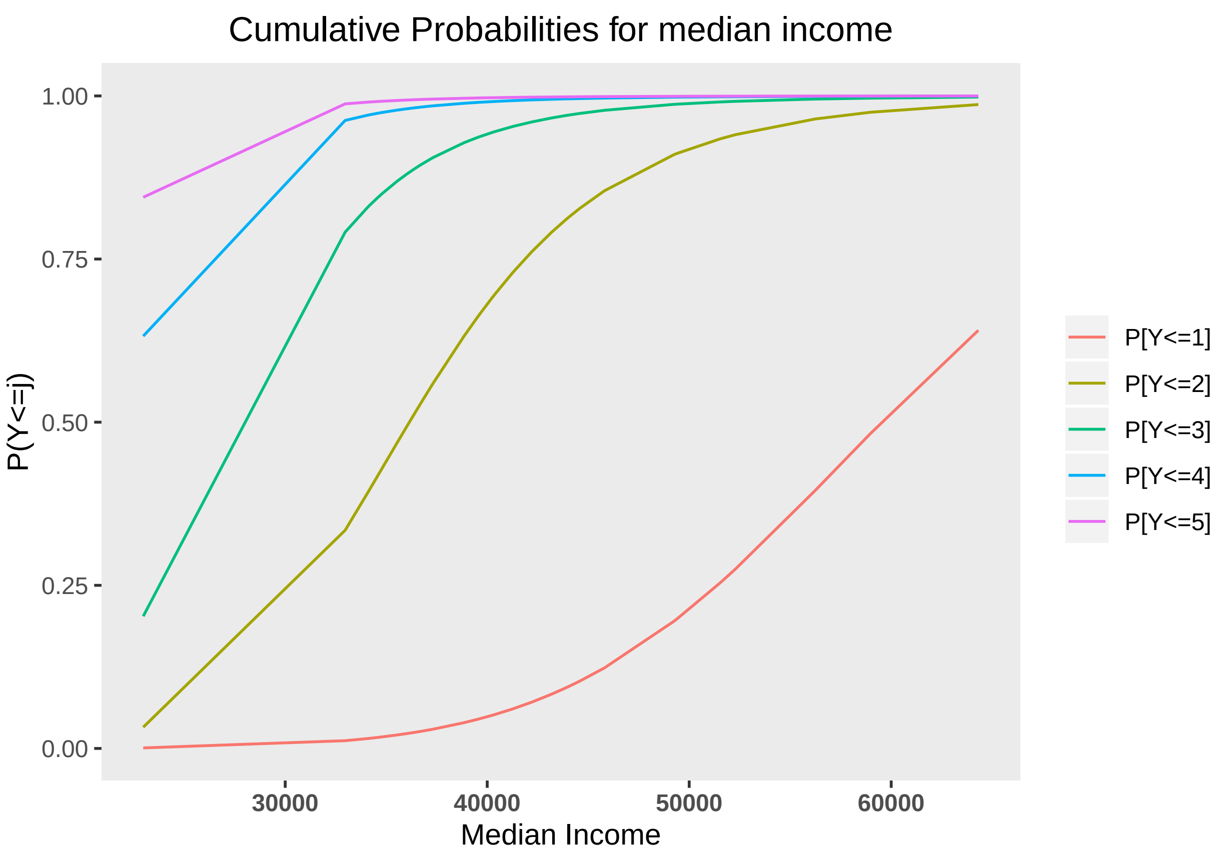

```
# dominance
dom_channel.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", median_income=43194.13,
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_dom <- predict(fit.select, newdata = dom_channel.test.ordnet1, type = "probs")
# plotting
classprob_dom_df = t(classprob_dom) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_dom_df) = NULL
colnames(classprob_dom_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_dom_df = as.data.frame(classprob_dom_df) %>%
  cbind(dom_channel.test.ordnet1) %>%
  dplyr::select(dominance, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         dominance = as.factor(dominance))

ggplot(classcumprob_dom_df, aes(x = dominance, y = probability)) +
```

```r
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Market Dominance",
         y = "P(Y<=j)",
         x= "Market Dominance") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



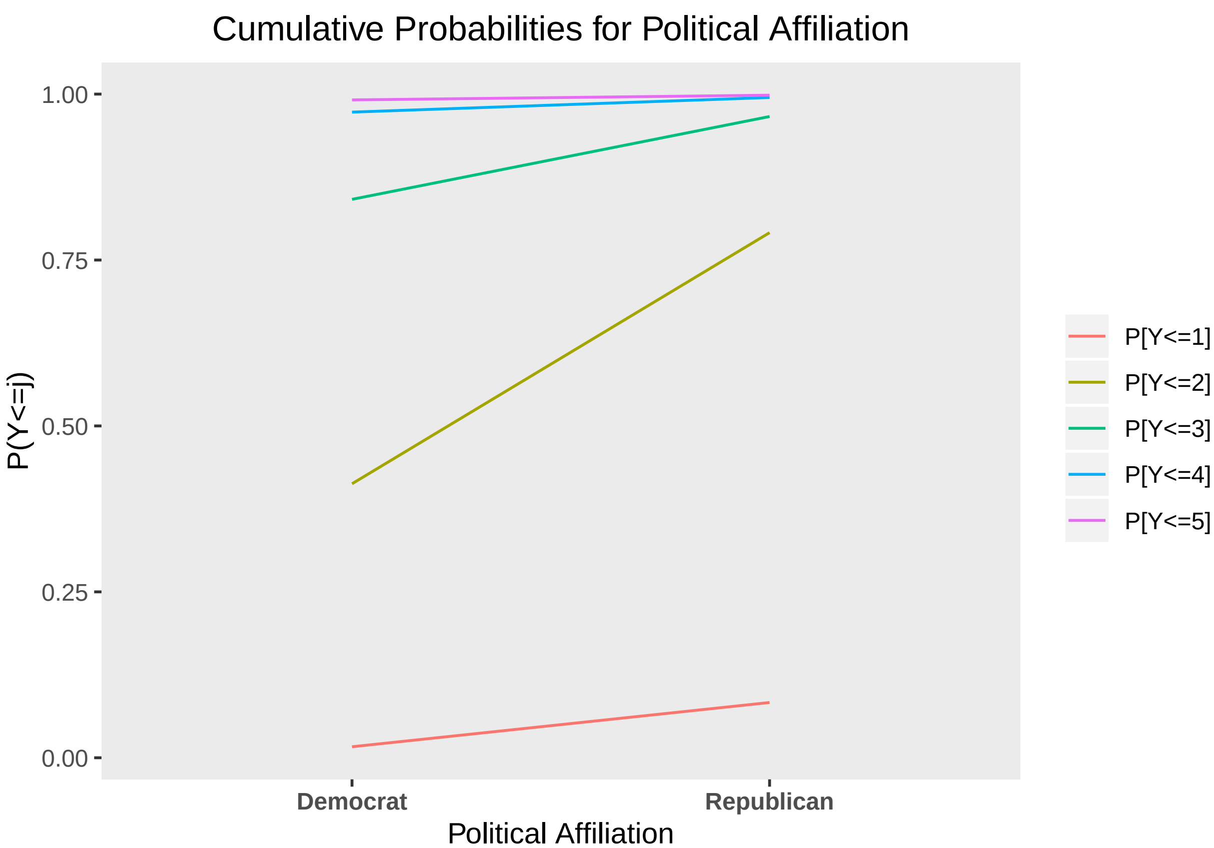

```r
# income
log_income.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_log_income <- predict(fit.select, newdata = log_income.test.ordnet1, type = "probs")
# plotting
classprob_log_income_df = t(classprob_log_income) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_log_income_df) = NULL
colnames(classprob_log_income_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_log_income_df = as.data.frame(classprob_log_income_df) %>%
  cbind(log_income.test.ordnet1) %>%
```

```
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_log_income_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for median income",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



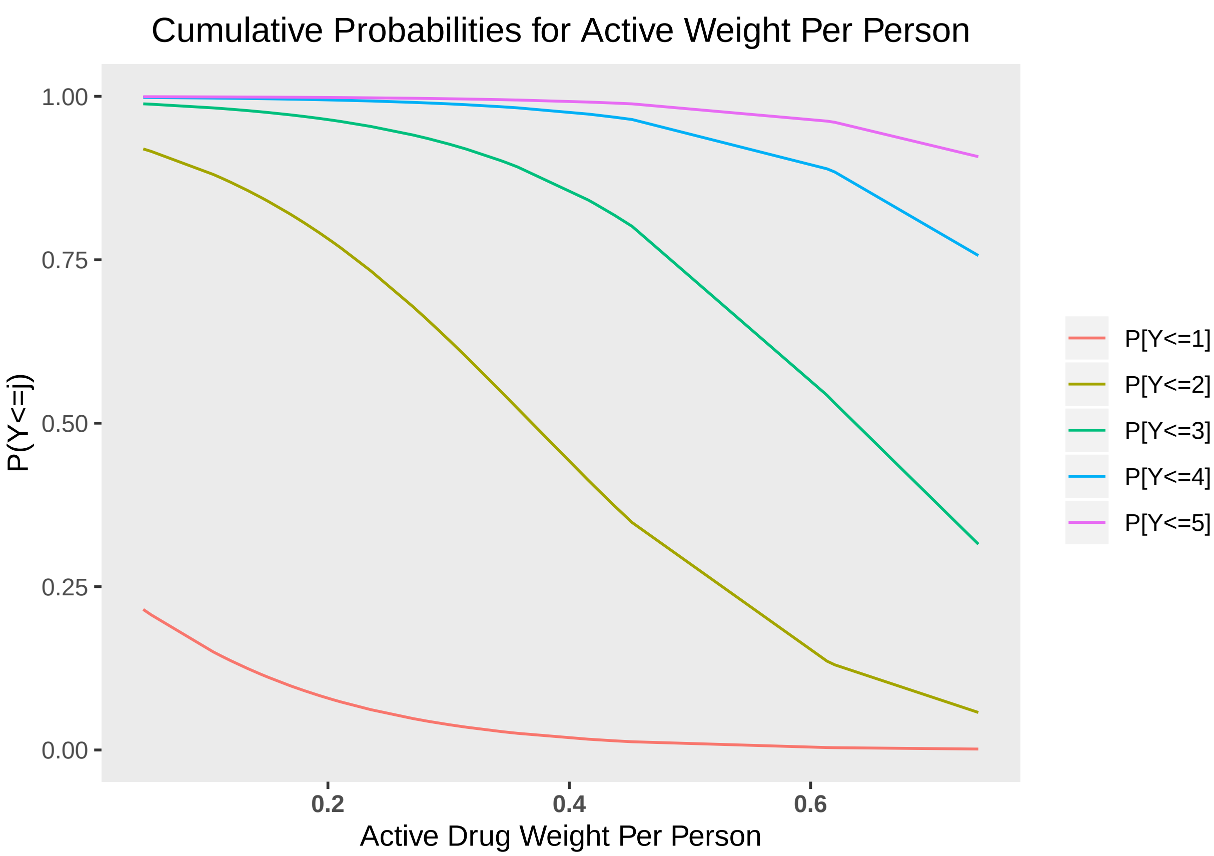

```
# political_aff
polaff.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
    mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_polaff <- predict(fit.select, newdata = polaff.test.ordnet1, type = "probs")
# plotting
classprob_polaff_df = t(classprob_polaff) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
```

```r
rownames(classprob_polaff_df) = NULL
colnames(classprob_polaff_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_polaff_df = as.data.frame(classprob_polaff_df) %>%
  cbind(polaff.test.ordnet1) %>%
  dplyr::select(political_aff, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         political_aff = as.factor(political_aff))

ggplot(classcumprob_polaff_df, aes(x = political_aff, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Political Affiliation",
         y = "P(Y<=j)",
         x= "Political Affiliation") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```

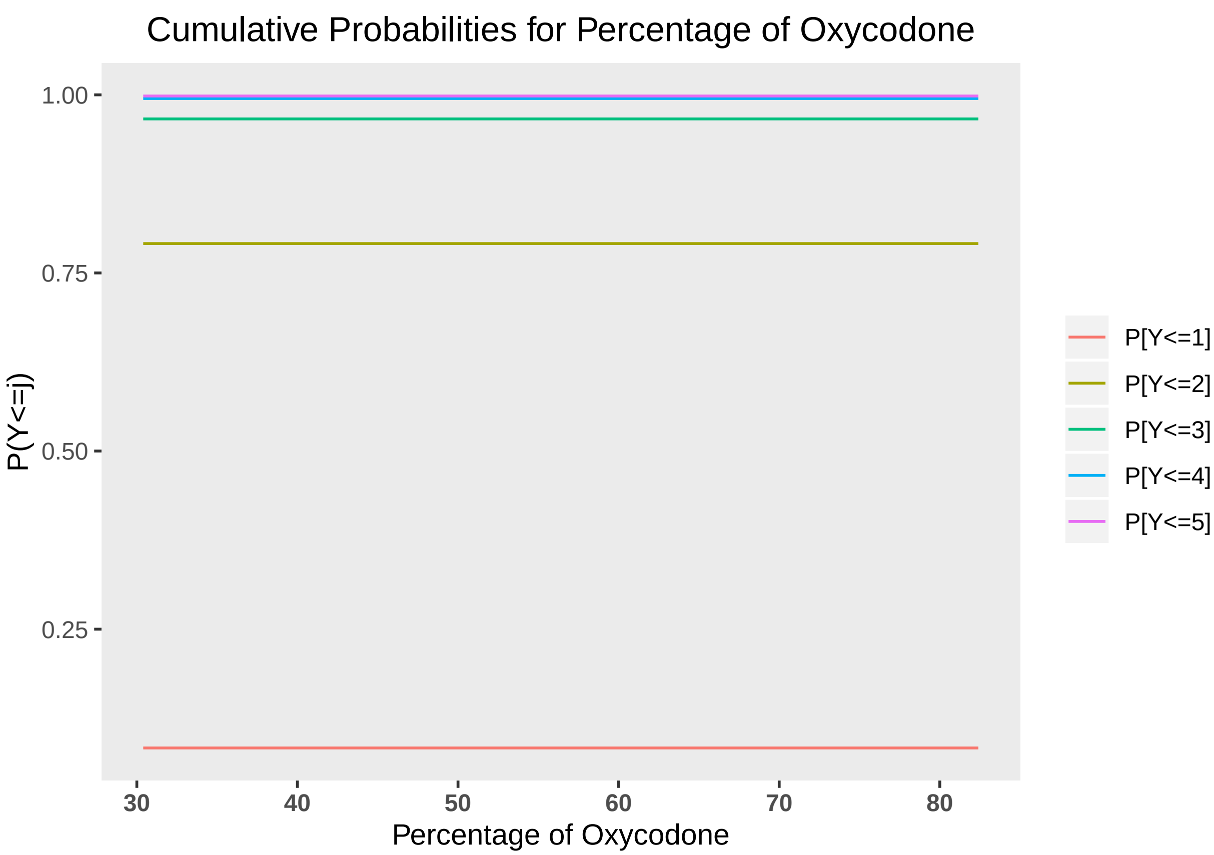



```r
# act_wt_person_county
act_wt.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         perc_oxy = 54.2, distr_num_ptt = 1.709)
```

```r
classprob_act_wt <- predict(fit.select, newdata = act_wt.test.ordnet1, type = "probs")
# plotting
classprob_act_wt_df = t(classprob_act_wt) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_act_wt_df) = NULL
colnames(classprob_act_wt_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_act_wt_df = as.data.frame(classprob_act_wt_df) %>%
  cbind(act_wt.test.ordnet1) %>%
  dplyr::select(act_wt_person_county, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_act_wt_df, aes(x = act_wt_person_county, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Active Weight Per Person",
         y = "P(Y<=j)",
         x= "Active Drug Weight Per Person") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```

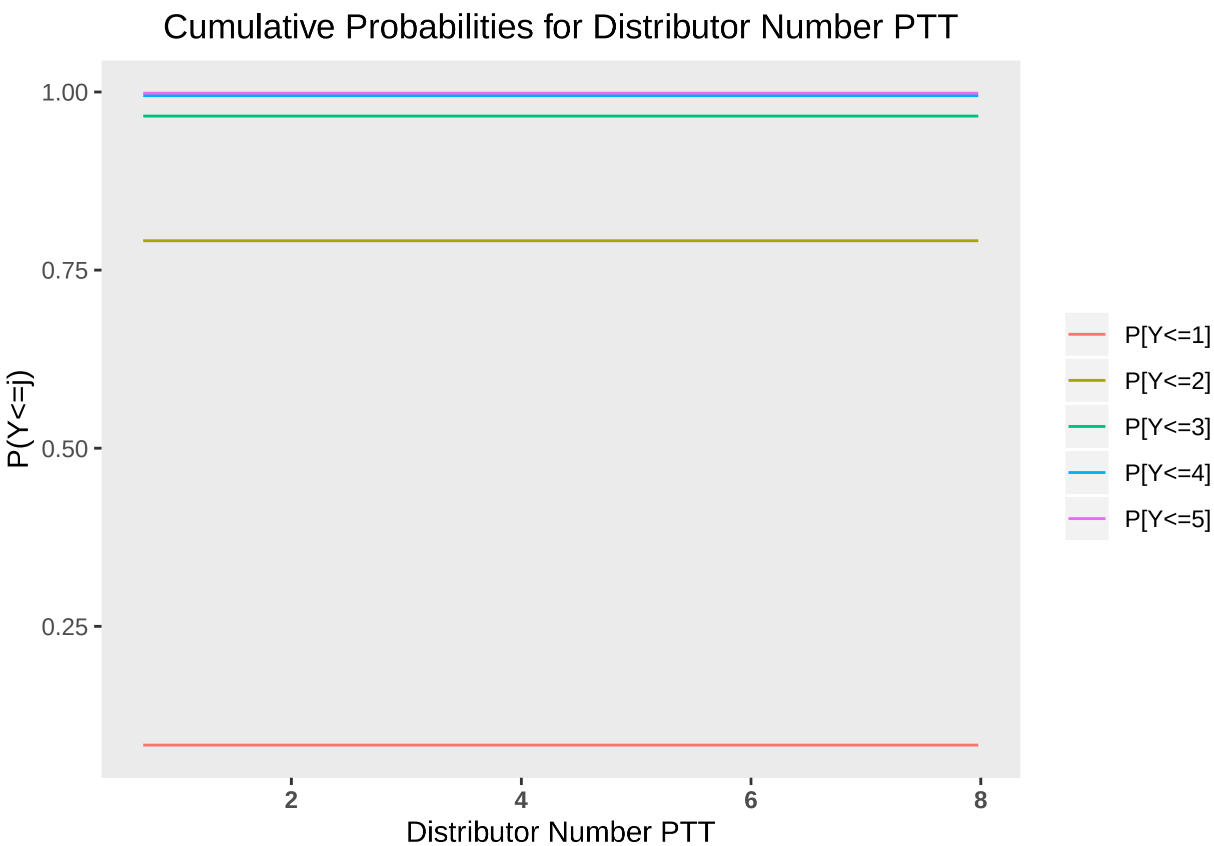


```r
# perc_oxy --> NOT SIGNIFICANT IN CURRENT MODEL
perc_oxy.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_perc_oxy <- predict(fit.select, newdata = perc_oxy.test.ordnet1, type = "probs")
# plotting
classprob_perc_oxy_df = t(classprob_perc_oxy) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_perc_oxy_df) = NULL
colnames(classprob_perc_oxy_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_perc_oxy_df = as.data.frame(classprob_perc_oxy_df) %>%
  cbind(perc_oxy.test.ordnet1) %>%
  dplyr::select(perc_oxy, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_perc_oxy_df, aes(x = perc_oxy, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Percentage of Oxycodone",
         y = "P(Y<=j)",
         x= "Percentage of Oxycodone") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```




```
# distr_num_ptt
distr_num.test.ordnet1 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         act_wt_person_county = 0.19294084)
classprob_distr_num  <- predict(fit.select, newdata = distr_num.test.ordnet1, type = "probs")
# plotting
classprob_distr_num_df = t(classprob_distr_num) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_distr_num_df) = NULL
colnames(classprob_distr_num_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_distr_num_df = as.data.frame(classprob_distr_num_df) %>%
  cbind(distr_num.test.ordnet1) %>%
  dplyr::select(distr_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_distr_num_df, aes(x = distr_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Distributor Number PTT",
         y = "P(Y<=j)",
         x= "Distributor Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
```

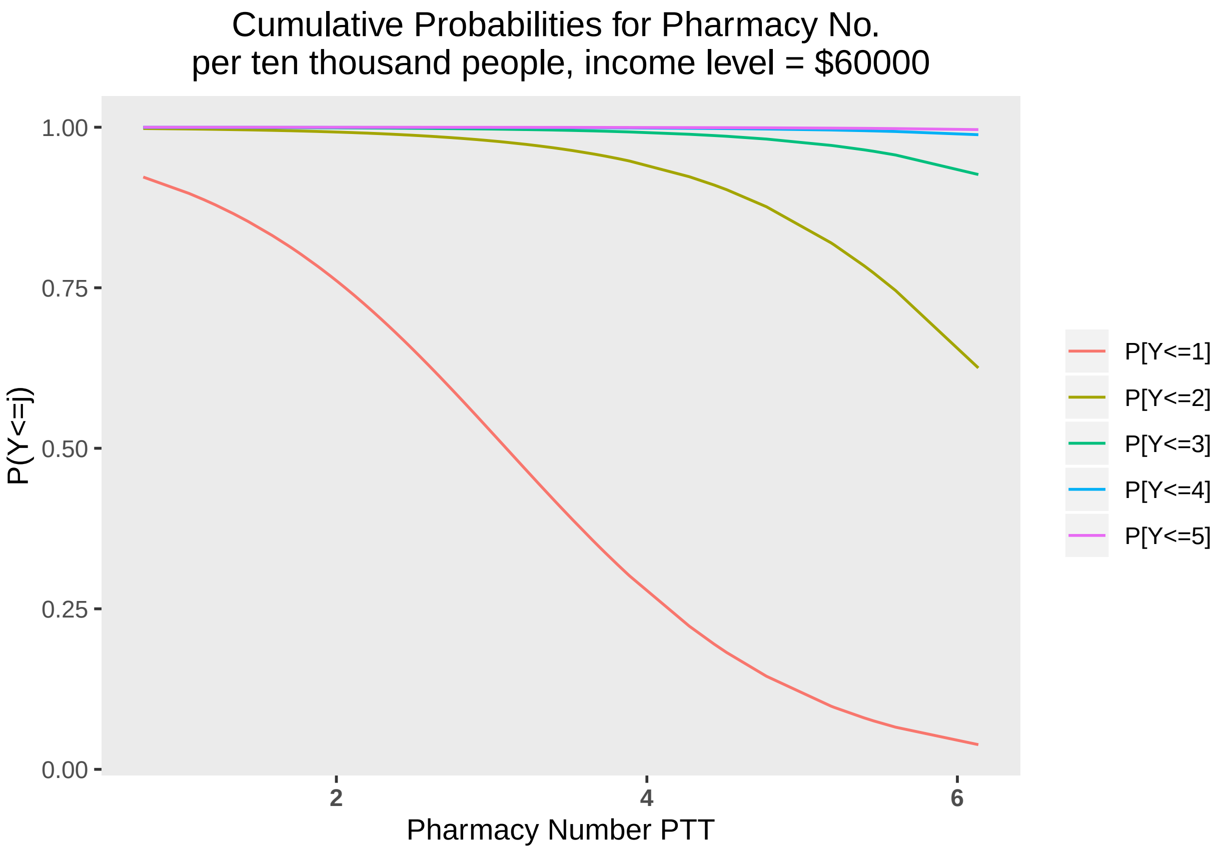
```
        axis.text.x = element_text(face = "bold"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank())
```



prediction plots for fit.select ON TRAINING DATA

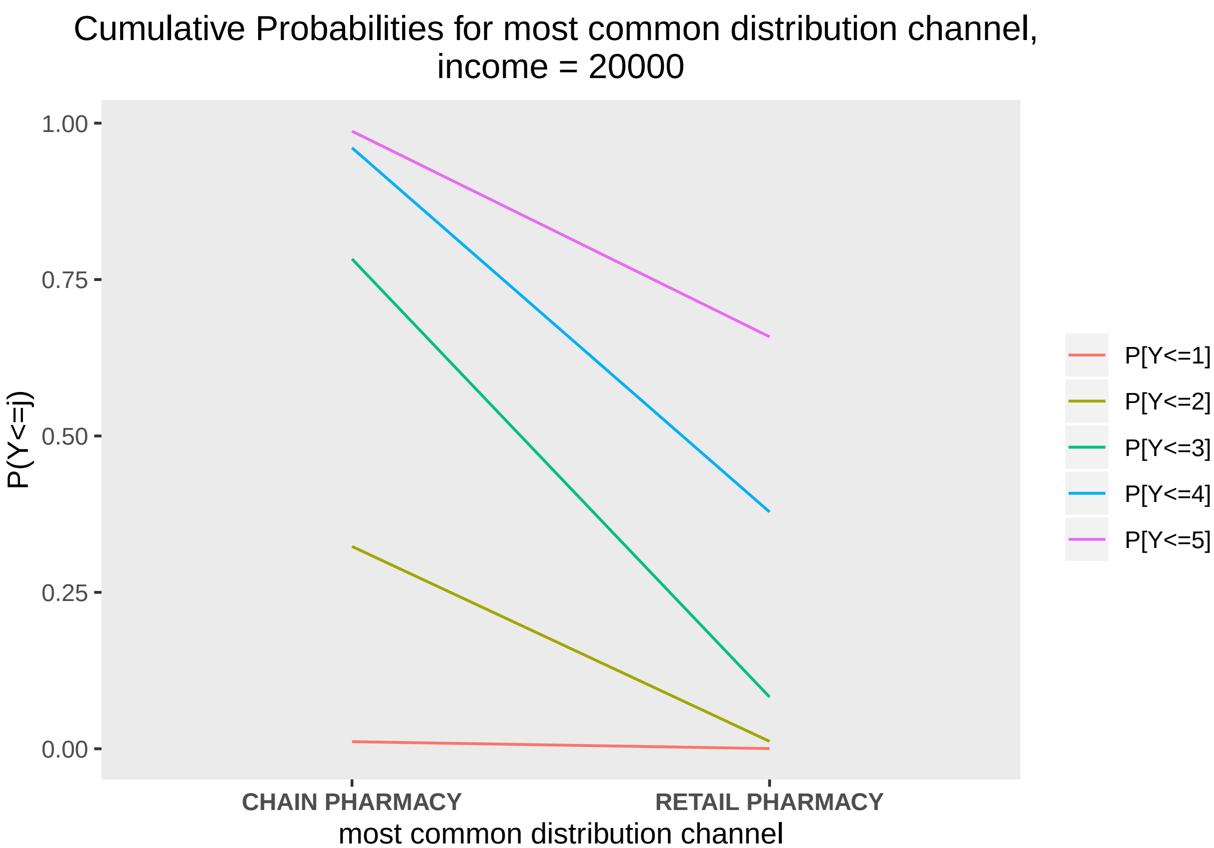
```
pred_matrix.train <- train_oh_wv_2012 %>%  #x is a matrix of predictors
  dplyr::select(est_death_rate_cat, pharmacy_num_ptt, most_dist_channel, dominance, median_income,
              political_aff, act_wt_person_county, distr_num_ptt)
#pharmacy no
pharm_num.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 60000,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_pharm_num  <- predict(fit.select, newdata = pharm_num.test.ordnet1, type = "probs") #class pr
# plotting
classprob_pharm_num_df = t(classprob_pharm_num) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_pharm_num_df) = NULL
colnames(classprob_pharm_num_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_pharm_num_df = as.data.frame(classprob_pharm_num_df) %>%
```

```
  cbind(pharm_num.test.ordnet1) %>%
  dplyr::select(pharmacy_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))
ggplot(classcumprob_pharm_num_df, aes(x = pharmacy_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Pharmacy No. \nper ten thousand people, income level = $6
         y = "P(Y<=j)",
         x= "Pharmacy Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



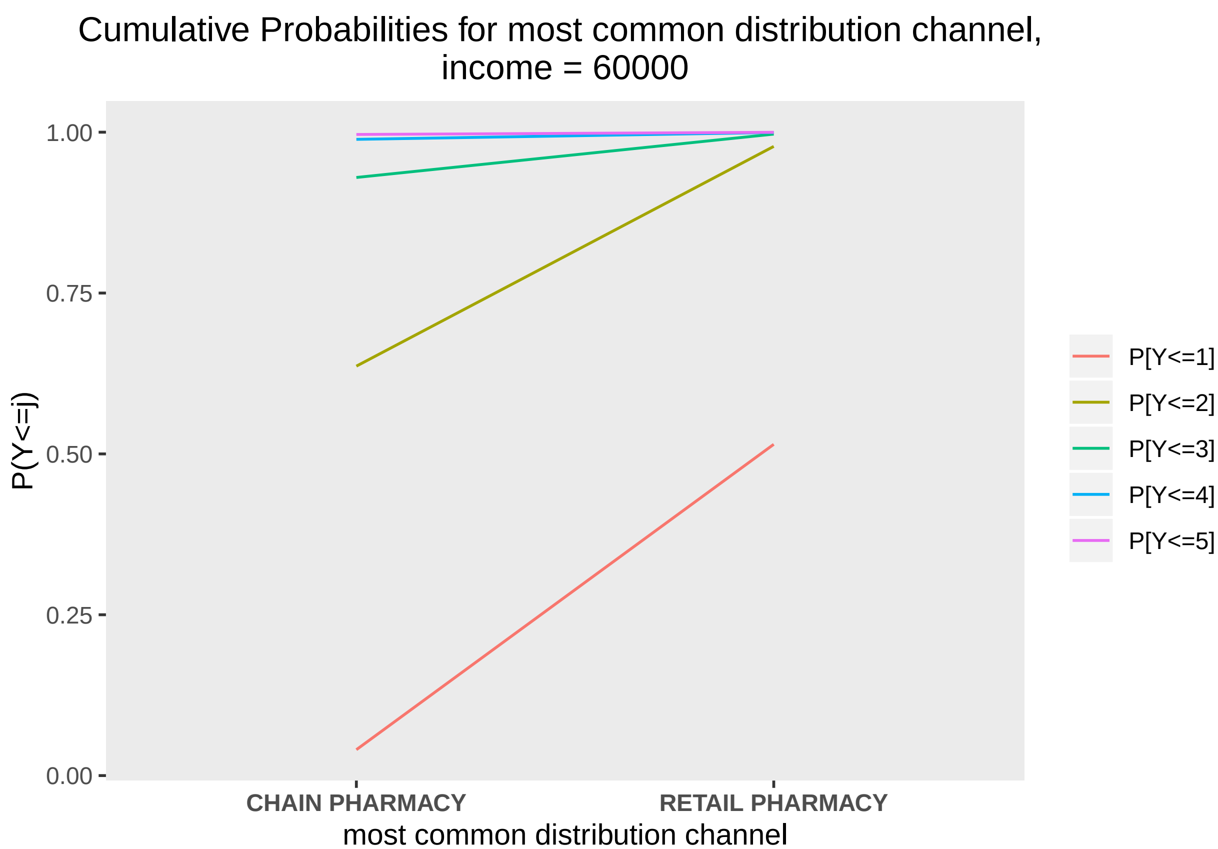

```
# most_dist_channel
pred_matrix.train <- train_oh_wv_2012 %>%  #x is a matrix of predictors
  dplyr::select(est_death_rate_cat, pharmacy_num_ptt, most_dist_channel, dominance, median_income,
              political_aff, act_wt_person_county, distr_num_ptt)
dom_channel.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
  mutate(pharmacy_num_ptt = 3.038, dominance="Yes", median_income=20000,
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_most_dist_channel <- predict(fit.select, newdata = dom_channel.test.ordnet1, type = "probs")
# plotting
classprob_most_dist_channel_df = t(classprob_most_dist_channel) %>%
```

```r
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_most_dist_channel_df) = NULL
colnames(classprob_most_dist_channel_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_most_dist_channel_df = as.data.frame(classprob_most_dist_channel_df) %>%
  cbind(dom_channel.test.ordnet1) %>%
  dplyr::select(most_dist_channel, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         most_dist_channel = as.factor(most_dist_channel))

ggplot(classcumprob_most_dist_channel_df, aes(x = most_dist_channel, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for most common distribution channel, \nincome = 20000",
         y = "P(Y<=j)",
         x= "most common distribution channel") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```




```r
#AT HIGHER INCOME LEVEL:
pred_matrix.train <- train_oh_wv_2012 %>%  #x is a matrix of predictors
  dplyr::select(est_death_rate_cat, pharmacy_num_ptt, most_dist_channel, dominance, median_income,
```

```
                political_aff, act_wt_person_county, distr_num_ptt)
dom_channel.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
  mutate(pharmacy_num_ptt = 3.038, dominance="Yes", median_income=60000,
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_most_dist_channel <- predict(fit.select, newdata = dom_channel.test.ordnet1, type = "probs")
# plotting
classprob_most_dist_channel_df = t(classprob_most_dist_channel) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_most_dist_channel_df) = NULL
colnames(classprob_most_dist_channel_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_most_dist_channel_df = as.data.frame(classprob_most_dist_channel_df) %>%
  cbind(dom_channel.test.ordnet1) %>%
  dplyr::select(most_dist_channel, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         most_dist_channel = as.factor(most_dist_channel))

ggplot(classcumprob_most_dist_channel_df, aes(x = most_dist_channel, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for most common distribution channel, \nincome = 60000",
         y = "P(Y<=j)",
         x= "most common distribution channel") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```

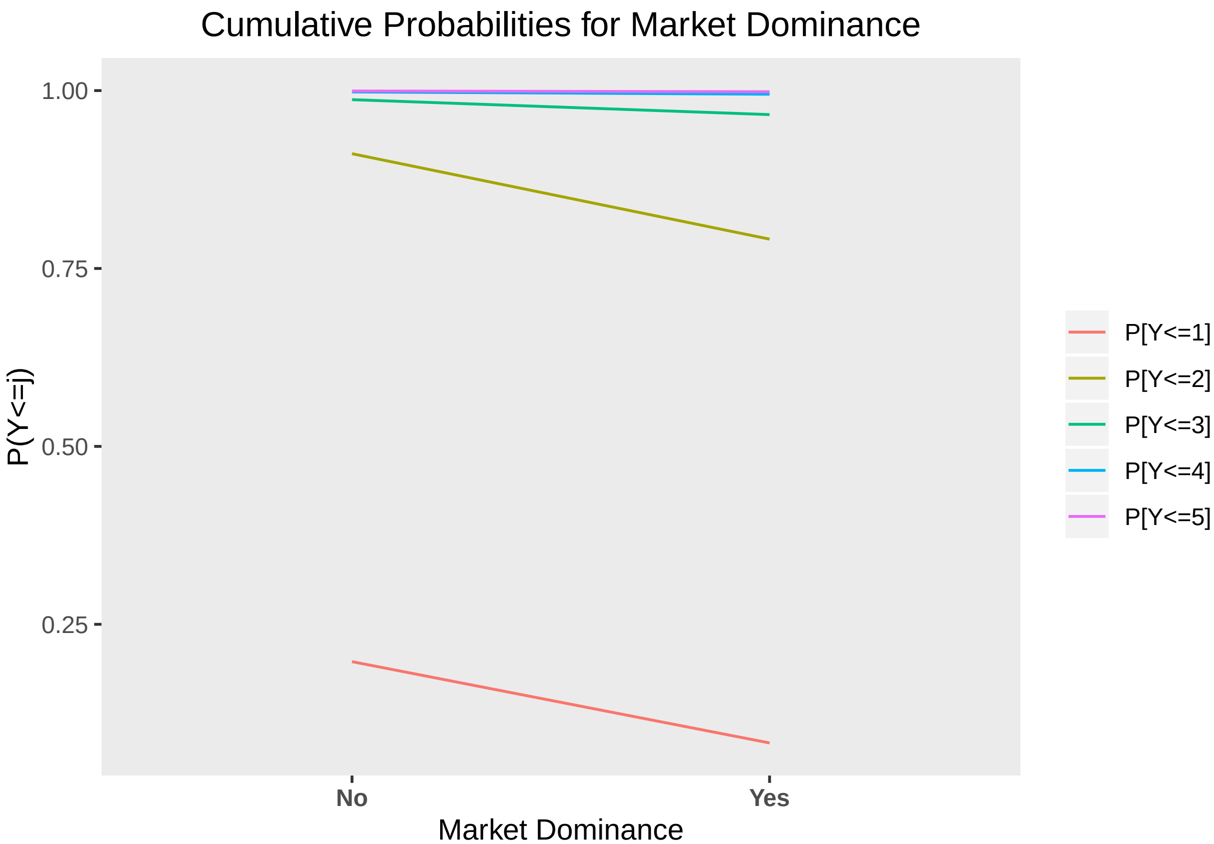



```r
# dominance
summary(fit.select)
```

```
## Call:
## polr(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     pharmacy_num_ptt:political_aff + most_dist_channel:log(median_income) +
##     pharmacy_num_ptt:log(median_income), data = train_oh_wv_2012,
##     Hess = TRUE, method = "logistic")
##
## Coefficients:
##                                                   Value Std. Error
## pharmacy_num_ptt                               -29.8071    13.1492
## most_dist_channelRETAIL PHARMACY                66.0704    27.7312
## dominanceYes                                     0.9971     0.5648
## log(median_income)                              -9.9542     3.4002
## political_affRepublican                          1.0636     1.3325
## act_wt_person_county                             7.5644     1.7395
## pharmacy_num_ptt:political_affRepublican        -0.9046     0.4408
## most_dist_channelRETAIL PHARMACY:log(median_income) -6.2989     2.6236
## pharmacy_num_ptt:log(median_income)              2.8877     1.2485
##                                               t value
## pharmacy_num_ptt                              -2.2668
## most_dist_channelRETAIL PHARMACY               2.3825
## dominanceYes                                   1.7653
## log(median_income)                            -2.9275
## political_affRepublican                        0.7982
```

```
## act_wt_person_county                                      4.3485
## pharmacy_num_ptt:political_affRepublican                 -2.0523
## most_dist_channelRETAIL PHARMACY:log(median_income) -2.4009
## pharmacy_num_ptt:log(median_income)                       2.3129
##
## Intercepts:
##             Value     Std. Error t value
## cat_1|cat_2 -105.9514   36.0762   -2.9369
## cat_2|cat_3 -102.2201   35.9160   -2.8461
## cat_3|cat_4 -100.1991   35.8594   -2.7942
## cat_4|cat_5  -98.2888   35.8476   -2.7419
## cat_5|cat_6  -97.1360   35.8385   -2.7104
##
## Residual Deviance: 236.3985
## AIC: 264.3985
```

```
dom_channel.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", median_income=43194.13,
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_dom <- predict(fit.select, newdata = dom_channel.test.ordnet1, type = "probs")
# plotting
classprob_dom_df = t(classprob_dom) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_dom_df) = NULL
colnames(classprob_dom_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_dom_df = as.data.frame(classprob_dom_df) %>%
  cbind(dom_channel.test.ordnet1) %>%
  dplyr::select(dominance, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         dominance = as.factor(dominance))

ggplot(classcumprob_dom_df, aes(x = dominance, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Market Dominance",
         y = "P(Y<=j)",
         x= "Market Dominance") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```

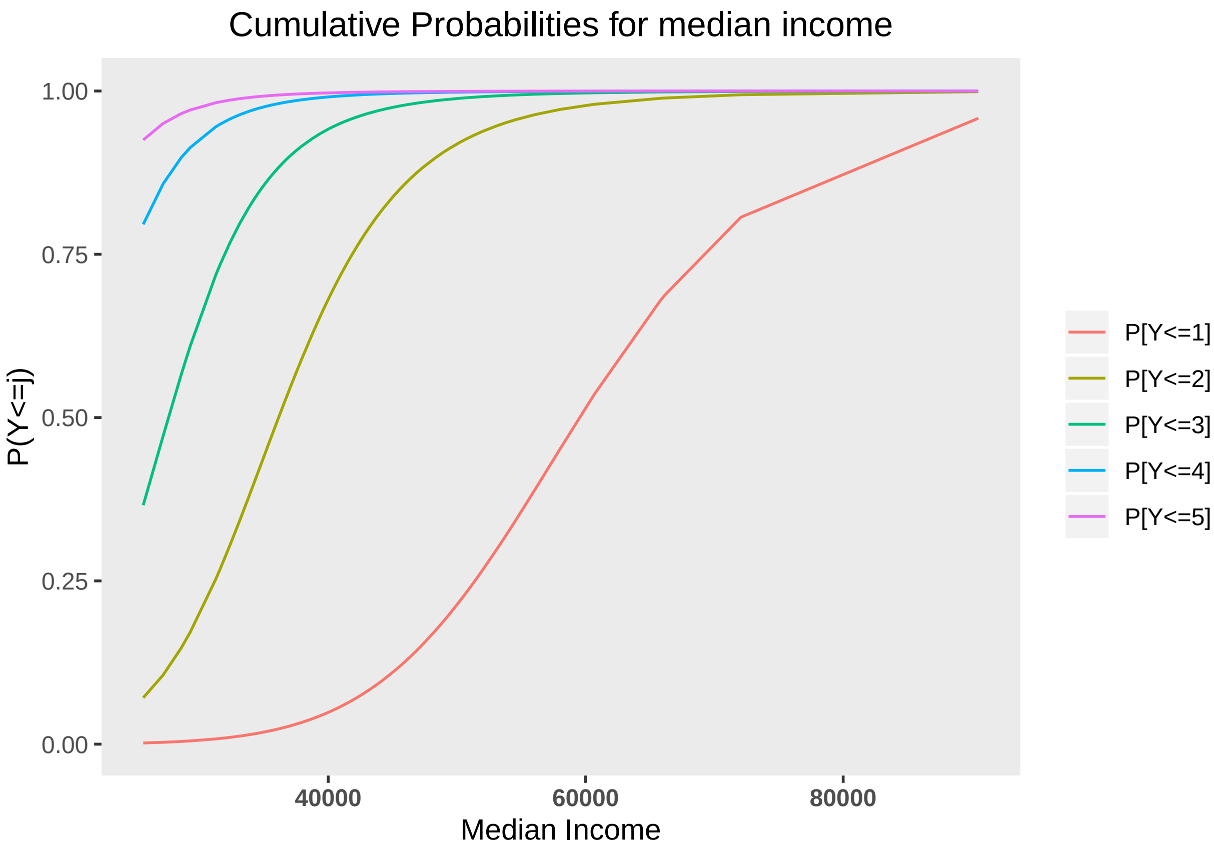



```r
# income
summary(fit.select)
```

```
## Call:
## polr(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     pharmacy_num_ptt:political_aff + most_dist_channel:log(median_income) +
##     pharmacy_num_ptt:log(median_income), data = train_oh_wv_2012,
##     Hess = TRUE, method = "logistic")
##
## Coefficients:
##                                                        Value Std. Error
## pharmacy_num_ptt                                    -29.8071    13.1492
## most_dist_channelRETAIL PHARMACY                     66.0704    27.7312
## dominanceYes                                          0.9971     0.5648
## log(median_income)                                   -9.9542     3.4002
## political_affRepublican                               1.0636     1.3325
## act_wt_person_county                                  7.5644     1.7395
## pharmacy_num_ptt:political_affRepublican             -0.9046     0.4408
## most_dist_channelRETAIL PHARMACY:log(median_income)  -6.2989     2.6236
## pharmacy_num_ptt:log(median_income)                   2.8877     1.2485
##                                                     t value
## pharmacy_num_ptt                                    -2.2668
## most_dist_channelRETAIL PHARMACY                     2.3825
## dominanceYes                                         1.7653
## log(median_income)                                  -2.9275
## political_affRepublican                              0.7982
```

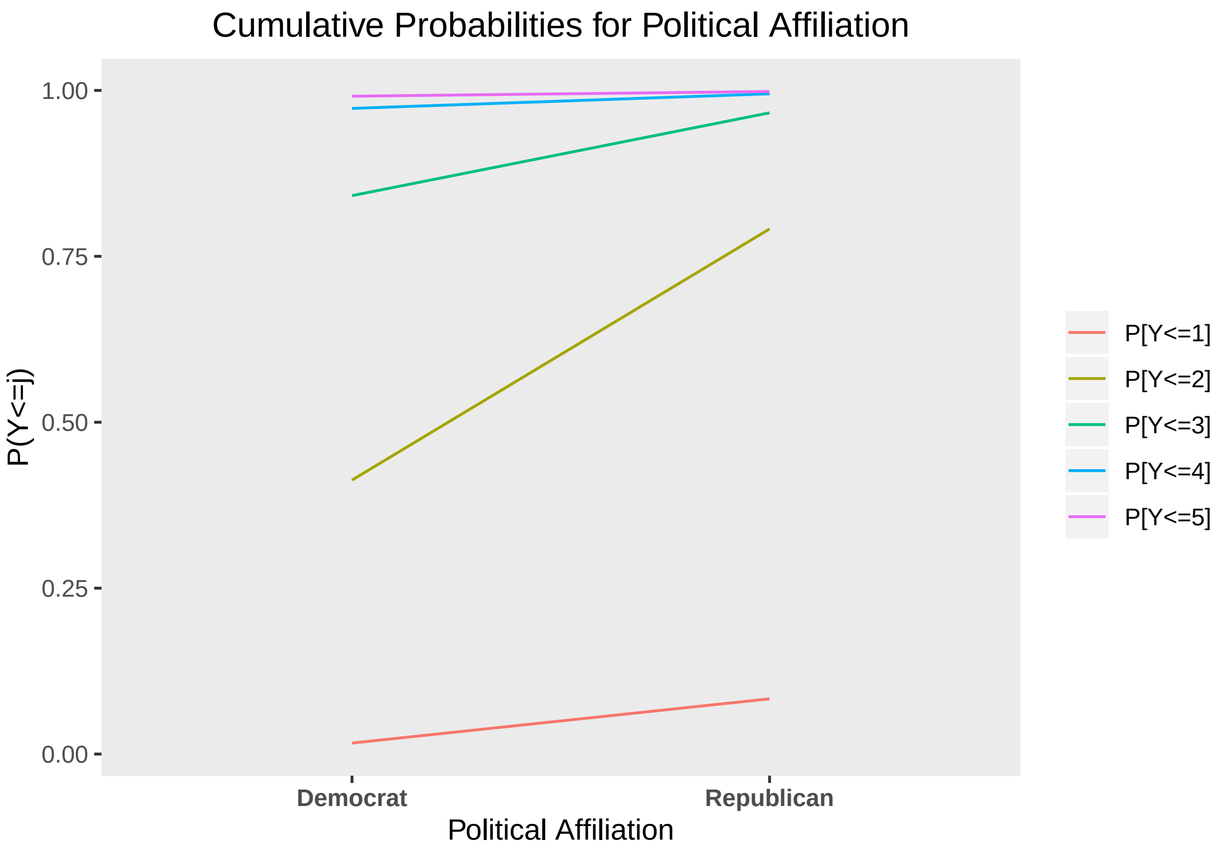
```
## act_wt_person_county                                     4.3485
## pharmacy_num_ptt:political_affRepublican                -2.0523
## most_dist_channelRETAIL PHARMACY:log(median_income) -2.4009
## pharmacy_num_ptt:log(median_income)                      2.3129
##
## Intercepts:
##             Value     Std. Error t value
## cat_1|cat_2 -105.9514  36.0762    -2.9369
## cat_2|cat_3 -102.2201  35.9160    -2.8461
## cat_3|cat_4 -100.1991  35.8594    -2.7942
## cat_4|cat_5  -98.2888  35.8476    -2.7419
## cat_5|cat_6  -97.1360  35.8385    -2.7104
##
## Residual Deviance: 236.3985
## AIC: 264.3985
```

```
log_income.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         political_aff = "Republican",
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_log_income <- predict(fit.select, newdata = log_income.test.ordnet1, type = "probs")
# plotting
classprob_log_income_df = t(classprob_log_income) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_log_income_df) = NULL
colnames(classprob_log_income_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_log_income_df = as.data.frame(classprob_log_income_df) %>%
  cbind(log_income.test.ordnet1) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_log_income_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for median income",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



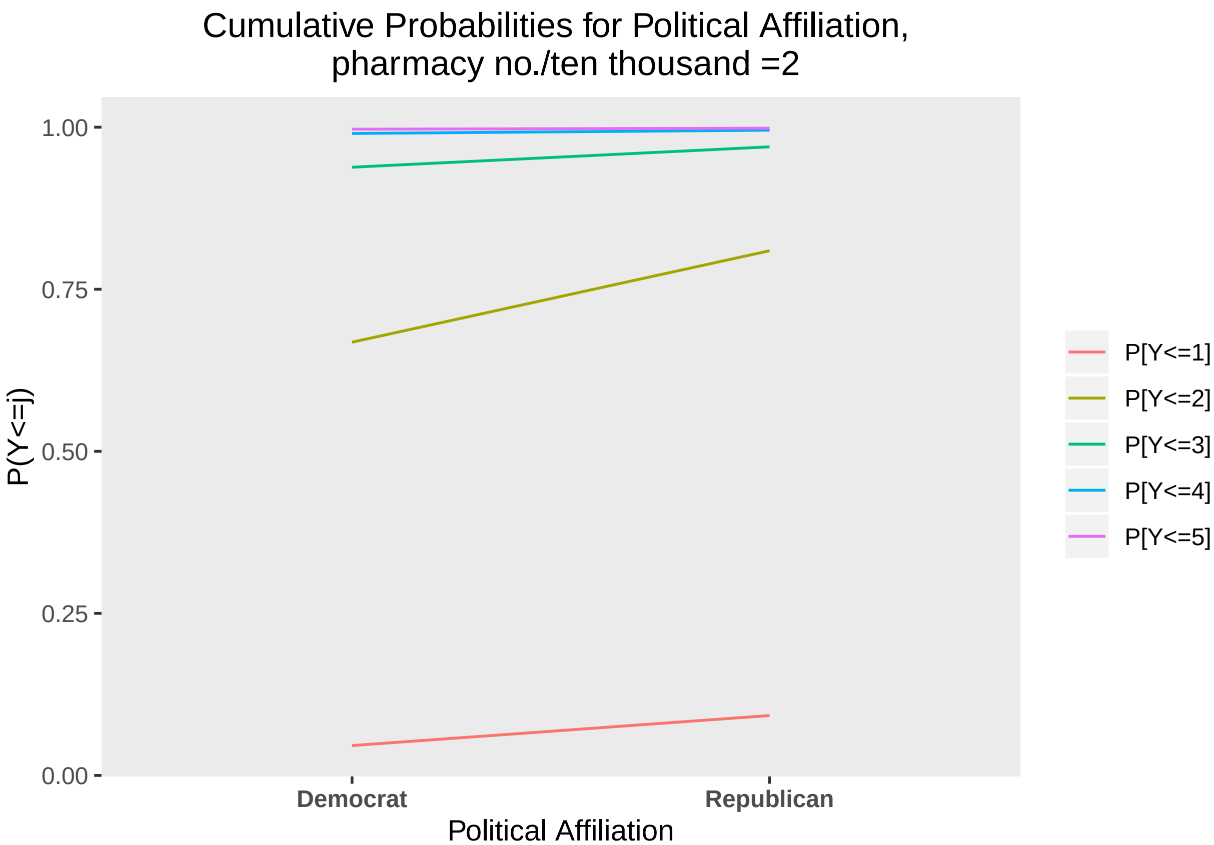

```r
# political_aff
polaff.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
    mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_polaff <- predict(fit.select, newdata = polaff.test.ordnet1, type = "probs")
# plotting
classprob_polaff_df = t(classprob_polaff) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_polaff_df) = NULL
colnames(classprob_polaff_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_polaff_df = as.data.frame(classprob_polaff_df) %>%
  cbind(polaff.test.ordnet1) %>%
  dplyr::select(political_aff, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         political_aff = as.factor(political_aff))

ggplot(classcumprob_polaff_df, aes(x = political_aff, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Political Affiliation",
         y = "P(Y<=j)",
         x= "Political Affiliation") +
```

```r
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



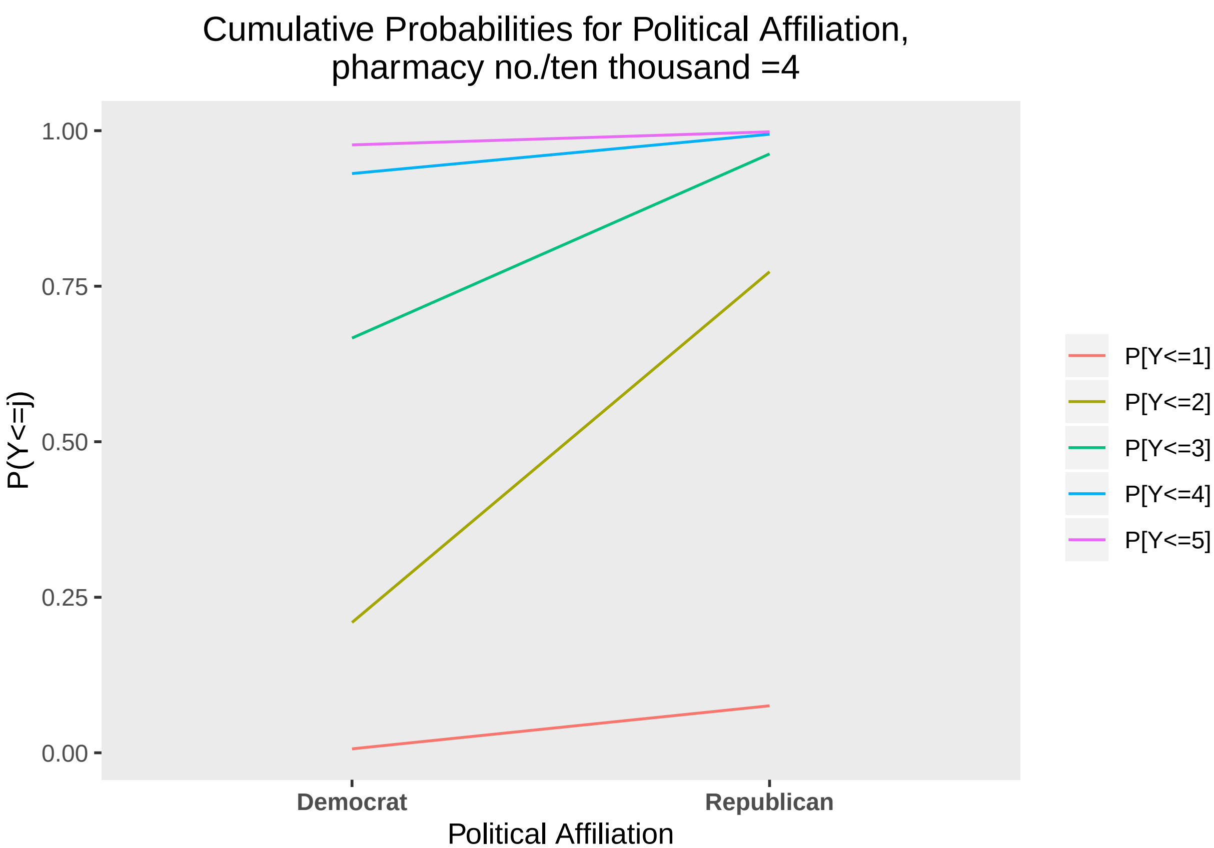

```r
# political_aff
polaff.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
    mutate(pharmacy_num_ptt = 2, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_polaff <- predict(fit.select, newdata = polaff.test.ordnet1, type = "probs")
# plotting
classprob_polaff_df = t(classprob_polaff) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_polaff_df) = NULL
colnames(classprob_polaff_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_polaff_df = as.data.frame(classprob_polaff_df) %>%
  cbind(polaff.test.ordnet1) %>%
  dplyr::select(political_aff, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         political_aff = as.factor(political_aff))
```

```r
ggplot(classcumprob_polaff_df, aes(x = political_aff, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Political Affiliation, \n pharmacy no./ten thousand =2",
         y = "P(Y<=j)",
         x= "Political Affiliation") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



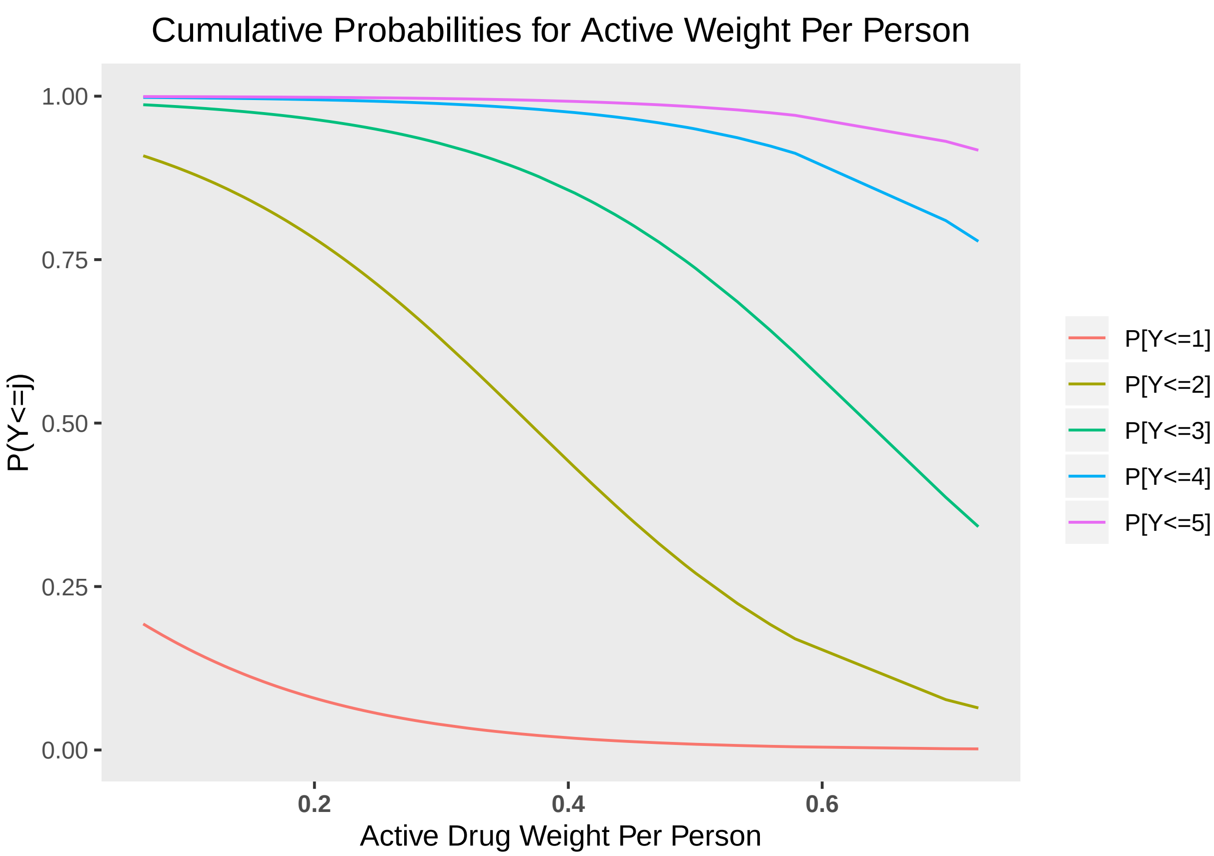

```r
#change pharm num to 4
polaff.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
    mutate(pharmacy_num_ptt = 4, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         act_wt_person_county = 0.19294084,
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_polaff <- predict(fit.select, newdata = polaff.test.ordnet1, type = "probs")
# plotting
classprob_polaff_df = t(classprob_polaff) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_polaff_df) = NULL
colnames(classprob_polaff_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
```

```r
classcumprob_polaff_df = as.data.frame(classprob_polaff_df) %>%
  cbind(polaff.test.ordnet1) %>%
  dplyr::select(political_aff, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class),
         political_aff = as.factor(political_aff))

ggplot(classcumprob_polaff_df, aes(x = political_aff, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Political Affiliation, \n pharmacy no./ten thousand =4",
         y = "P(Y<=j)",
         x= "Political Affiliation") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



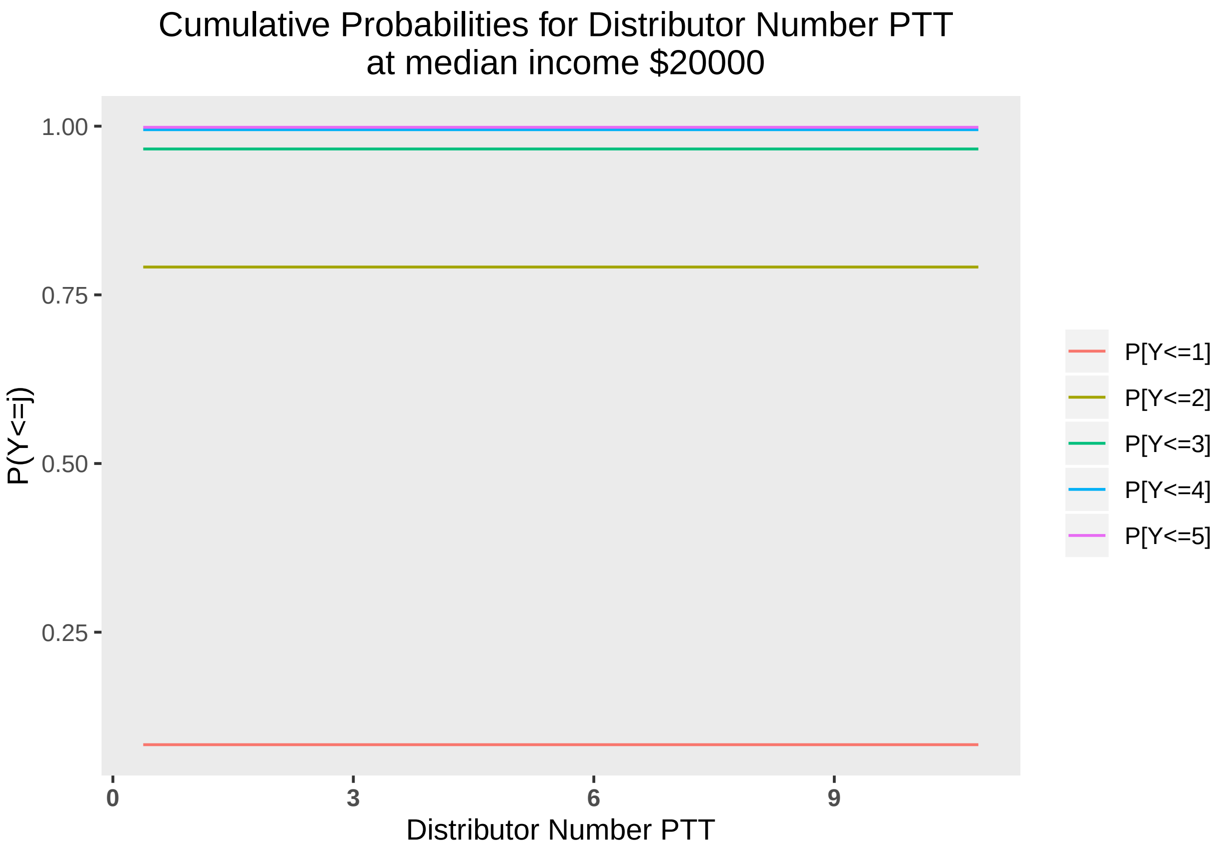

```r
# act_wt_person_county
act_wt.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         perc_oxy = 54.2, distr_num_ptt = 1.709)
classprob_act_wt <- predict(fit.select, newdata = act_wt.test.ordnet1, type = "probs")
# plotting
classprob_act_wt_df = t(classprob_act_wt) %>%
```

```
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_act_wt_df) = NULL
colnames(classprob_act_wt_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_act_wt_df = as.data.frame(classprob_act_wt_df) %>%
  cbind(act_wt.test.ordnet1) %>%
  dplyr::select(act_wt_person_county, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_act_wt_df, aes(x = act_wt_person_county, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Active Weight Per Person",
         y = "P(Y<=j)",
         x= "Active Drug Weight Per Person") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```




```
# distr_num_ptt
distr_num.test.ordnet1 = as.data.frame(pred_matrix.train) %>%
  mutate(pharmacy_num_ptt = 3.038, most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
```

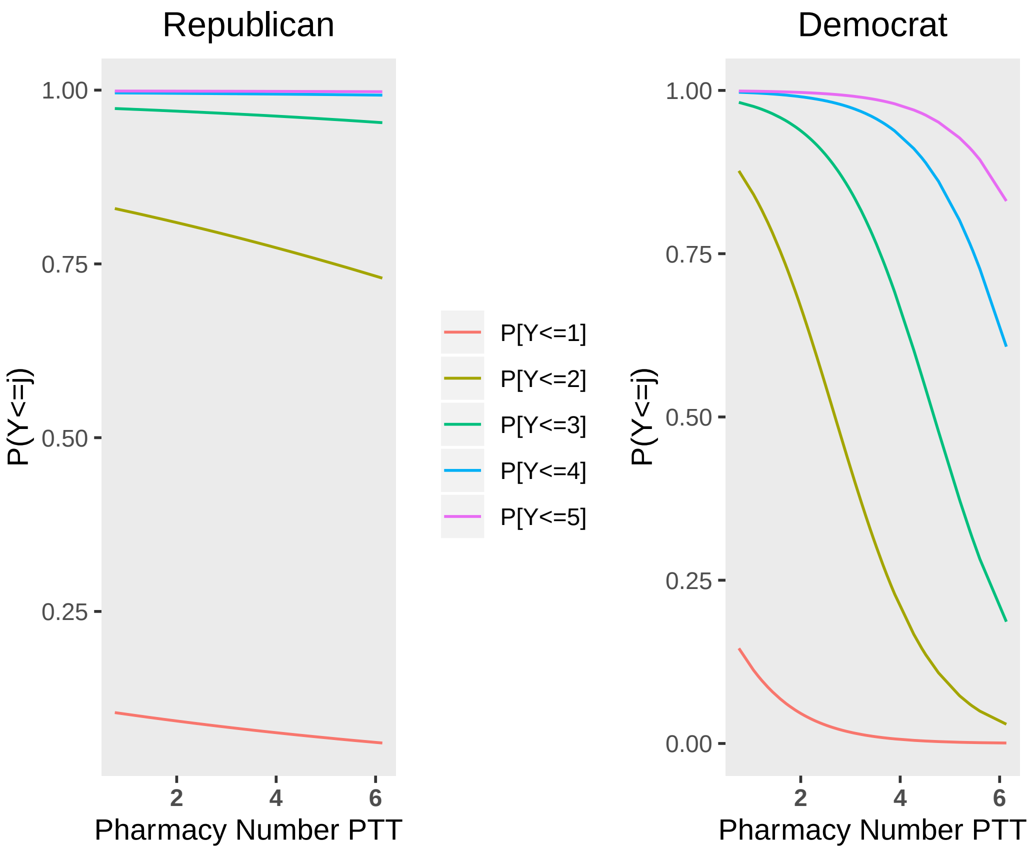
```
         political_aff="Republican",
         act_wt_person_county = 0.19294084)
classprob_distr_num  <- predict(fit.select, newdata = distr_num.test.ordnet1, type = "probs")
# plotting
classprob_distr_num_df = t(classprob_distr_num) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_distr_num_df) = NULL
colnames(classprob_distr_num_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_distr_num_df = as.data.frame(classprob_distr_num_df) %>%
  cbind(distr_num.test.ordnet1) %>%
  dplyr::select(distr_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

ggplot(classcumprob_distr_num_df, aes(x = distr_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Cumulative Probabilities for Distributor Number PTT \n at median income $20000",
         y = "P(Y<=j)",
         x= "Distributor Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
```



interaction plots for fit.select ON TRAINING DATA

```
## number of pharmacies and political affilition
# Republican
pharm_num.test.ordnet1_rep = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_pharm_num_rep  <- predict(fit.select, newdata = pharm_num.test.ordnet1_rep, type = "probs", s
# plotting
classprob_pharm_num_rep_df = t(classprob_pharm_num_rep) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_pharm_num_rep_df) = NULL
colnames(classprob_pharm_num_rep_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_pharm_num_rep_df = as.data.frame(classprob_pharm_num_rep_df) %>%
  cbind(pharm_num.test.ordnet1_rep) %>%
  dplyr::select(pharmacy_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

num_pharm_rep_p = ggplot(classcumprob_pharm_num_rep_df, aes(x = pharmacy_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
```

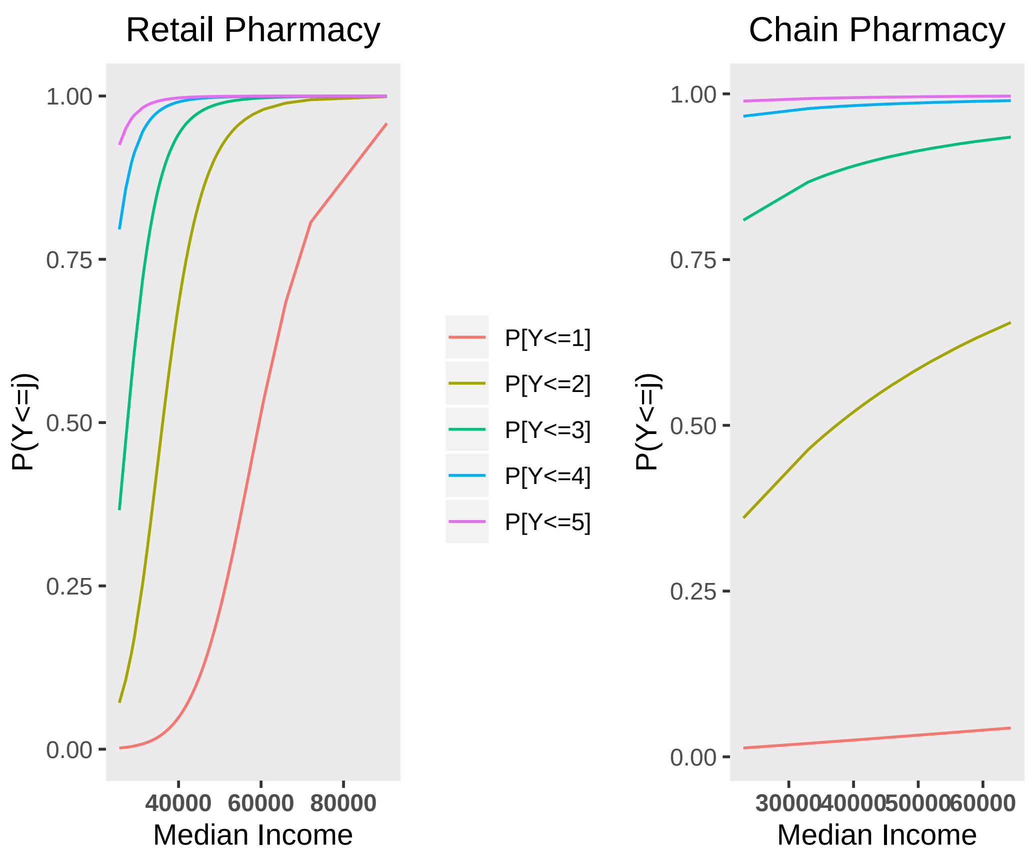
```r
    labs(title = "Republican",
         y = "P(Y<=j)",
         x= "Pharmacy Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
# Democrat
pharm_num.test.ordnet1_dem = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         median_income = 43194.13,
         political_aff="Democrat",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_pharm_num_dem  <- predict(fit.select, newdata = pharm_num.test.ordnet1_dem, type = "probs", s
# plotting
classprob_pharm_num_dem_df = t(classprob_pharm_num_dem) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_pharm_num_dem_df) = NULL
colnames(classprob_pharm_num_dem_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_pharm_num_dem_df = as.data.frame(classprob_pharm_num_dem_df) %>%
  cbind(pharm_num.test.ordnet1_dem) %>%
  dplyr::select(pharmacy_num_ptt, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

num_pharm_dem_p = ggplot(classcumprob_pharm_num_dem_df, aes(x = pharmacy_num_ptt, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Democrat",
         y = "P(Y<=j)",
         x= "Pharmacy Number PTT") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          legend.position = "none")
#num_pharm_rep_p + num_pharm_dem_p
plot_grid(num_pharm_rep_p, num_pharm_dem_p, axis = "r", align = "v")
```




```
# probability of falling below a category decreases faster for democrats. is
# significant? because our dataset is small.
```

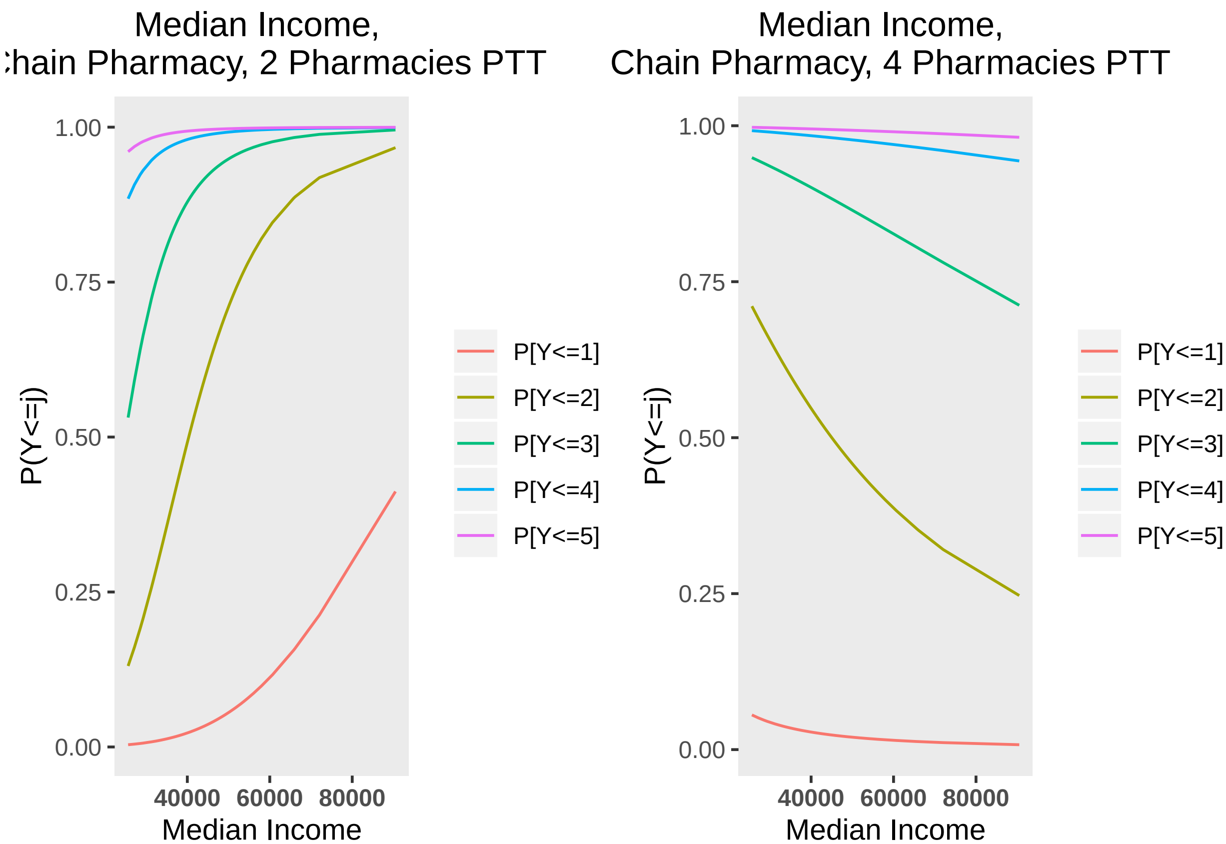
```
## distribution channel and income
# RETAIL PHARMACRY
income.test.ordnet1_rp = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 3.038,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_rp  <- predict(fit.select, newdata = income.test.ordnet1_rp, type = "probs", se.fit = 
# plotting
classprob_income_rp_df = t(classprob_income_rp) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_rp_df) = NULL
colnames(classprob_income_rp_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_rp_df = as.data.frame(classprob_income_rp_df) %>%
  cbind(income.test.ordnet1_rp) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_rp_p = ggplot(classcumprob_income_rp_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
```

```r
    labs(title = "Retail Pharmacy",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
# CHAIN PHARMACRY
income.test.ordnet1_cp = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "CHAIN PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 3.038,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_cp  <- predict(fit.select, newdata = income.test.ordnet1_cp, type = "probs", se.fit = T
# plotting
classprob_income_cp_df = t(classprob_income_cp) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_cp_df) = NULL
colnames(classprob_income_cp_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_cp_df = as.data.frame(classprob_income_cp_df) %>%
  cbind(income.test.ordnet1_cp) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_cp_p = ggplot(classcumprob_income_cp_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Chain Pharmacy",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          legend.position = "none")
plot_grid(income_rp_p, income_cp_p, axis = "r", align = "v")
```

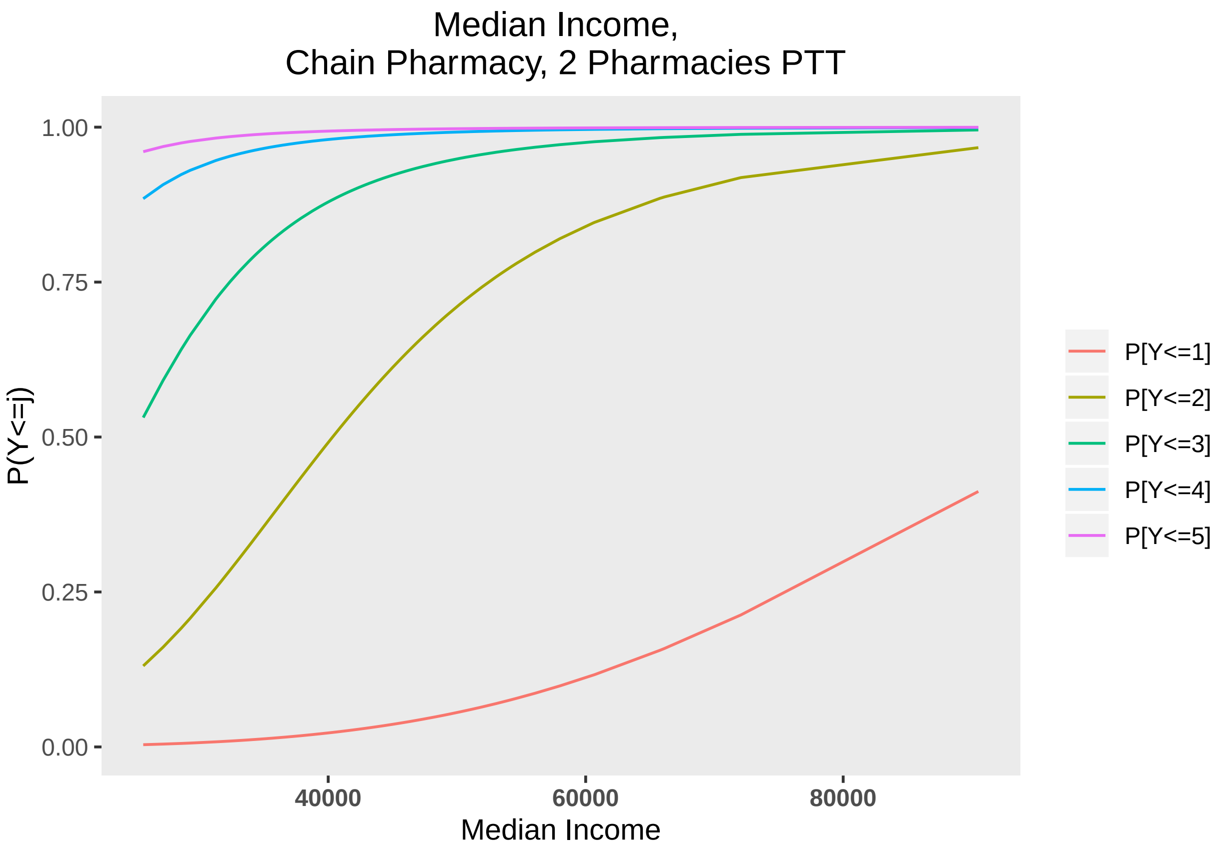



```r
#pharm num at 2 and 4 for chain
income.test.ordnet1_cp = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "CHAIN PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 2,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_cp  <- predict(fit.select, newdata = income.test.ordnet1_cp, type = "probs", se.fit =
# plotting
classprob_income_cp_df = t(classprob_income_cp) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_cp_df) = NULL
colnames(classprob_income_cp_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_cp_df = as.data.frame(classprob_income_cp_df) %>%
  cbind(income.test.ordnet1_cp) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_cp_p.pharm2 = ggplot(classcumprob_income_cp_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Median Income, \n Chain Pharmacy, 2 Pharmacies PTT",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
```

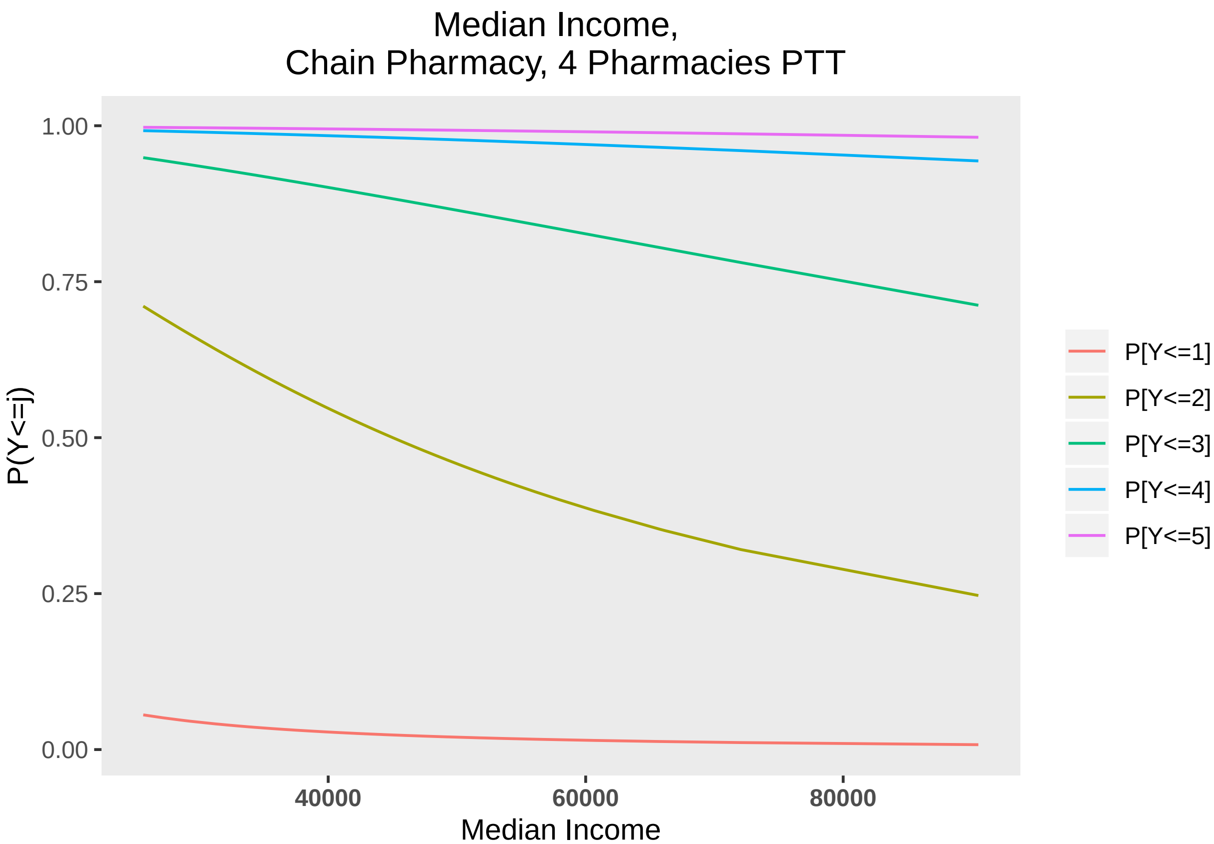
```
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
income.test.ordnet1_cp = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "CHAIN PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 4,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_cp  <- predict(fit.select, newdata = income.test.ordnet1_cp, type = "probs", se.fit = 
# plotting
classprob_income_cp_df = t(classprob_income_cp) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_cp_df) = NULL
colnames(classprob_income_cp_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_cp_df = as.data.frame(classprob_income_cp_df) %>%
  cbind(income.test.ordnet1_cp) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_cp_p.pharm4 = ggplot(classcumprob_income_cp_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Median Income, \n Chain Pharmacy, 4 Pharmacies PTT",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
plot_grid(income_cp_p.pharm2, income_cp_p.pharm4)
```




```
income_cp_p.pharm2
```

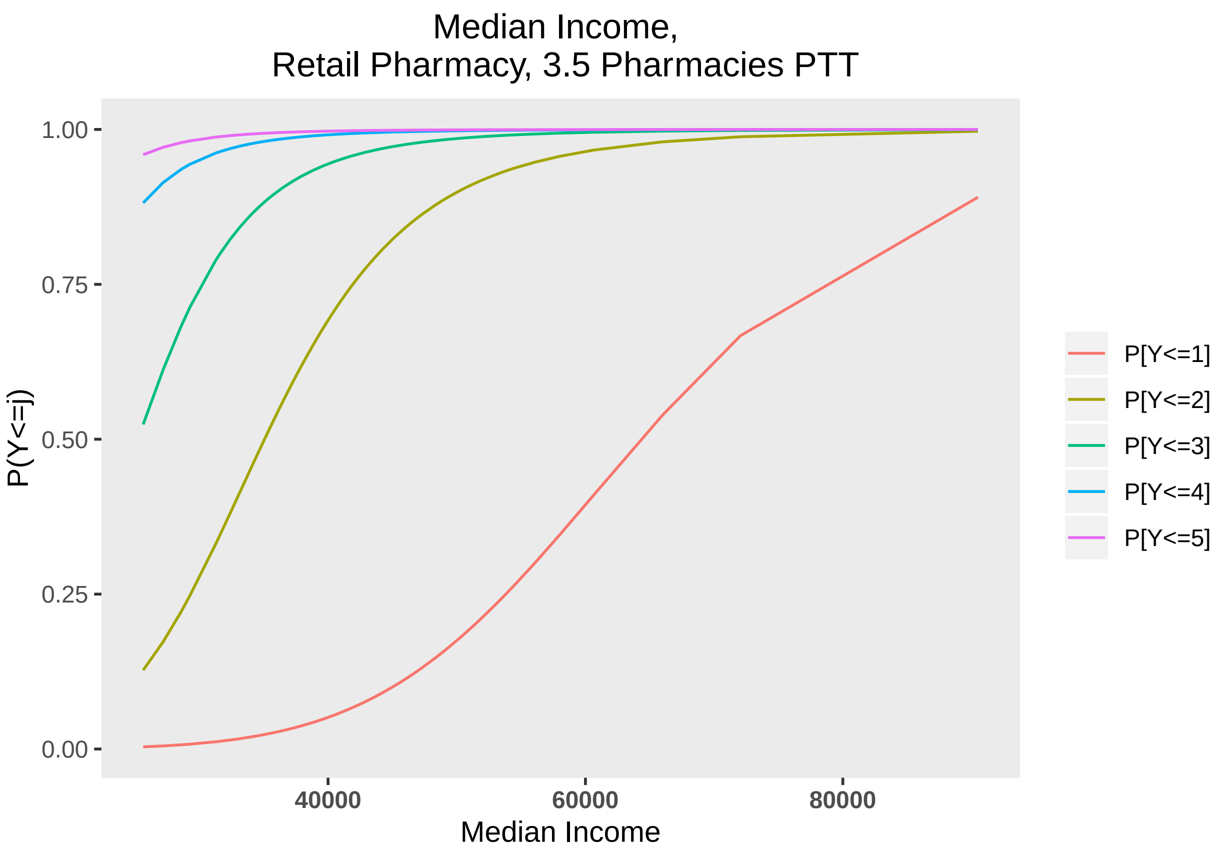



```
income_cp_p.pharm4
```



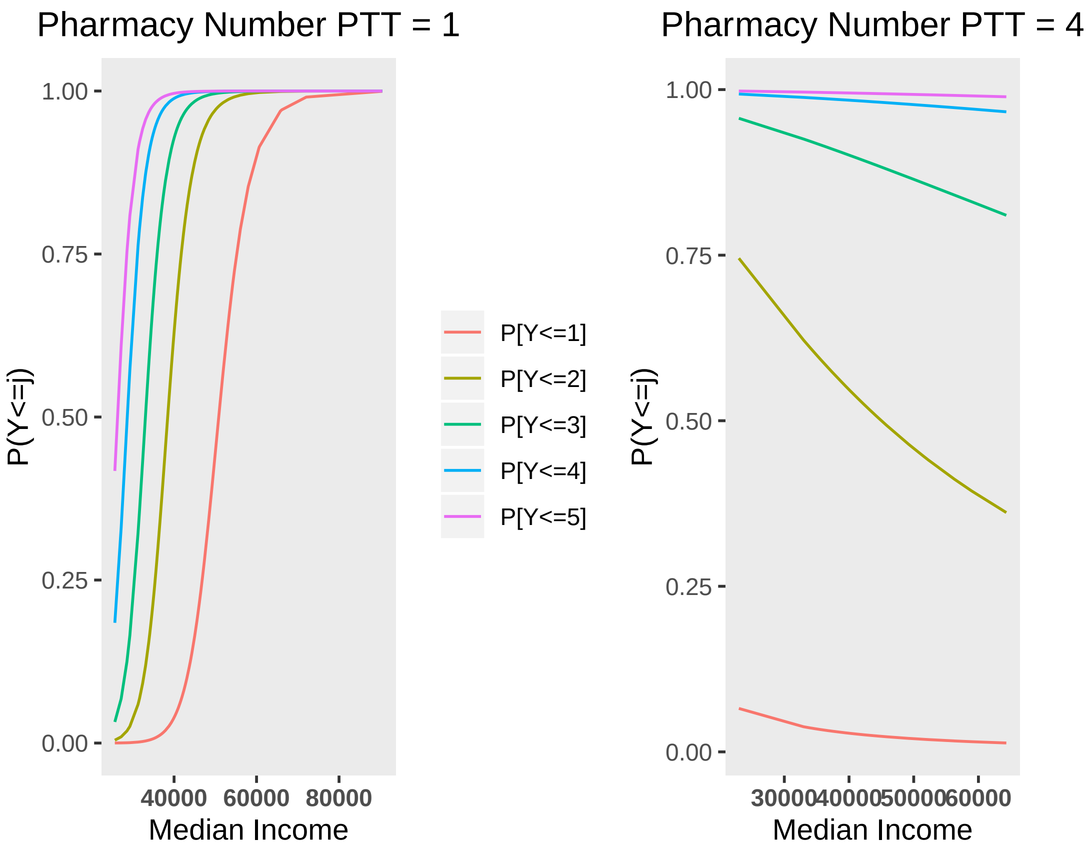

```
income.test.ordnet1_cp = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 3.5,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_cp  <- predict(fit.select, newdata = income.test.ordnet1_cp, type = "probs", se.fit = 
# plotting
classprob_income_cp_df = t(classprob_income_cp) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_cp_df) = NULL
colnames(classprob_income_cp_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_cp_df = as.data.frame(classprob_income_cp_df) %>%
  cbind(income.test.ordnet1_cp) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_cp_p.retail.3.5 = ggplot(classcumprob_income_cp_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Median Income, \n Retail Pharmacy, 3.5 Pharmacies PTT",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
```

```
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
income_cp_p.retail.3.5
```




```
## pharmacy number and log median income
# range of pharmacy number ptt: 0.928 4.850
# fixing pharmacy number at = 1
income.test.ordnet1_1 = as.data.frame(pred_matrix.train) %>%
  mutate(most_dist_channel = "RETAIL PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 1,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_1  <- predict(fit.select, newdata = income.test.ordnet1_1, type = "probs", se.fit = TRU
# plotting
classprob_income_1_df = t(classprob_income_1) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_1_df) = NULL
colnames(classprob_income_1_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_1_df = as.data.frame(classprob_income_1_df) %>%
  cbind(income.test.ordnet1_1) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))
```

```r
income_1_p = ggplot(classcumprob_income_1_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Pharmacy Number PTT = 1",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank())
# fixing pharmacy number at = 4
income.test.ordnet1_4 = as.data.frame(pred_matrix.ordnet1) %>%
  mutate(most_dist_channel = "CHAIN PHARMACY", dominance = "Yes",
         pharmacy_num_ptt = 4,
         political_aff="Republican",
         perc_oxy = 54.2, act_wt_person_county = 0.19294084, distr_num_ptt = 1.709)
classprob_income_4  <- predict(fit.select, newdata = income.test.ordnet1_4, type = "probs", se.fit = TRU
# plotting
classprob_income_4_df = t(classprob_income_4) %>%
  as.data.frame() %>%
  cumsum() %>% t() %>% as.data.frame() %>%
  dplyr::select(-`cat_6`)
rownames(classprob_income_4_df) = NULL
colnames(classprob_income_4_df) = c("P[Y<=1]", "P[Y<=2]", "P[Y<=3]", "P[Y<=4]", "P[Y<=5]")
# plotting
classcumprob_income_4_df = as.data.frame(classprob_income_4_df) %>%
  cbind(income.test.ordnet1_4) %>%
  dplyr::select(median_income, `P[Y<=1]`:`P[Y<=5]`) %>%
  gather(key = "class", value = "probability", `P[Y<=1]`:`P[Y<=5]`) %>%
  mutate(class = as.factor(class))

income_4_p = ggplot(classcumprob_income_4_df, aes(x = median_income, y = probability)) +
    geom_line(aes(color = class, group = class)) +
    labs(title = "Pharmacy Number PTT = 4",
         y = "P(Y<=j)",
         x= "Median Income") +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(face = "bold"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          legend.position = "none")
#income_1_p + income_4_p
plot_grid(income_1_p, income_4_p, axis = "r", align = "v")
```



##Check proportional odds assumption

```
summary(fit.select)
```

```
## Call:
## polr(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
##     dominance + log(median_income) + political_aff + act_wt_person_county +
##     pharmacy_num_ptt:political_aff + most_dist_channel:log(median_income) +
##     pharmacy_num_ptt:log(median_income), data = train_oh_wv_2012,
##     Hess = TRUE, method = "logistic")
##
## Coefficients:
##                                                          Value Std. Error
## pharmacy_num_ptt                                      -29.8071    13.1492
## most_dist_channelRETAIL PHARMACY                       66.0704    27.7312
## dominanceYes                                            0.9971     0.5648
## log(median_income)                                     -9.9542     3.4002
## political_affRepublican                                 1.0636     1.3325
## act_wt_person_county                                    7.5644     1.7395
## pharmacy_num_ptt:political_affRepublican               -0.9046     0.4408
## most_dist_channelRETAIL PHARMACY:log(median_income)   -6.2989     2.6236
## pharmacy_num_ptt:log(median_income)                     2.8877     1.2485
##                                                       t value
## pharmacy_num_ptt                                      -2.2668
## most_dist_channelRETAIL PHARMACY                       2.3825
## dominanceYes                                           1.7653
## log(median_income)                                    -2.9275
## political_affRepublican                                0.7982
```

```
## act_wt_person_county                                  4.3485
## pharmacy_num_ptt:political_affRepublican            -2.0523
## most_dist_channelRETAIL PHARMACY:log(median_income) -2.4009
## pharmacy_num_ptt:log(median_income)                  2.3129
##
## Intercepts:
##             Value     Std. Error t value
## cat_1|cat_2 -105.9514   36.0762    -2.9369
## cat_2|cat_3 -102.2201   35.9160    -2.8461
## cat_3|cat_4 -100.1991   35.8594    -2.7942
## cat_4|cat_5  -98.2888   35.8476    -2.7419
## cat_5|cat_6  -97.1360   35.8385    -2.7104
##
## Residual Deviance: 236.3985
## AIC: 264.3985
```

```r
library(ordinal)
```

```
##
## Attaching package: 'ordinal'

## The following objects are masked from 'package:brms':
##
##     ranef, VarCorr

## The following object is masked from 'package:dplyr':
##
##     slice
```

```r
clm<-clm(formula = est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel +
    dominance + log(median_income) + political_aff + act_wt_person_county +
    pharmacy_num_ptt:political_aff + most_dist_channel:log(median_income) +
    pharmacy_num_ptt:log(median_income), data=train_oh_wv_2012,
      link = c("logit"))
```

```
## Warning: (3) Model is nearly unidentifiable: large eigenvalue ratio
##  - Rescale variables?
## In addition: Absolute and relative convergence criteria were met
```

```r
nominal_test(clm) #no evidence of non-proportional odds
```

```
## Tests of nominal effects
##
## formula: est_death_rate_cat ~ pharmacy_num_ptt + most_dist_channel + dominance + log(median_income) +
##                                     Df  logLik    AIC    LRT Pr(>Chi)
## <none>                                 -118.19 264.37
## pharmacy_num_ptt                     4 -115.78 267.57 4.8010   0.3083
## most_dist_channel                    4 -117.14 270.28 2.0881   0.7196
## dominance                            4 -115.15 266.30 6.0672   0.1942
## log(median_income)
## political_aff                        4 -115.60 267.19 5.1759   0.2697
## act_wt_person_county
## pharmacy_num_ptt:political_aff
## most_dist_channel:log(median_income)
## pharmacy_num_ptt:log(median_income)   4 -115.74 267.48 4.8930   0.2984
```