

# FinalProject

XinyiYu

2025-10-16

packages

```
library(readxl)
library(dplyr)

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(readr)
library(stringr)
library(ggplot2)
```

CDC Pregnancy Vaccination

```
#Use Webscraping or an API for Data Collection
vacc_url <- "https://data.cdc.gov/api/views/h7pm-wmjc/rows.csv?accessType=DOWNLOAD"
vacc_df <- read.csv(vacc_url)

#uniform NY
vacc_df$Geography <- gsub("NY-City of New York", "New York", vacc_df$Geography)
vacc_df$Geography <- gsub("NY-Rest of state", "New York", vacc_df$Geography)

#eliminate others
vacc_df <- vacc_df %>%
  filter(!(Geography %in% c("United States", "Puerto Rico", "Commonwealth of the Northern Mariana Islands")))

#State name to abbreviation
state_abbrev <- c(
  'Alabama' = 'AL', 'Alaska' = 'AK', 'Arizona' = 'AZ', 'Arkansas' = 'AR',
  'California' = 'CA', 'Colorado' = 'CO', 'Connecticut' = 'CT', 'Delaware' = 'DE',
  'District of Columbia' = 'DC', 'Florida' = 'FL', 'Georgia' = 'GA', 'Hawaii' = 'HI',
  'Idaho' = 'ID', 'Illinois' = 'IL', 'Indiana' = 'IN', 'Iowa' = 'IA', 'Kansas' = 'KS',
```

```

'Kentucky' = 'KY', 'Louisiana' = 'LA', 'Maine' = 'ME', 'Maryland' = 'MD',
'Massachusetts' = 'MA', 'Michigan' = 'MI', 'Minnesota' = 'MN',
'Mississippi' = 'MS', 'Missouri' = 'MO', 'Montana' = 'MT', 'Nebraska' = 'NE',
'Nevada' = 'NV', 'New Hampshire' = 'NH', 'New Jersey' = 'NJ',
'New Mexico' = 'NM', 'New York' = 'NY', 'North Carolina' = 'NC',
'North Dakota' = 'ND', 'Ohio' = 'OH', 'Oklahoma' = 'OK', 'Oregon' = 'OR',
'Pennsylvania' = 'PA', 'Rhode Island' = 'RI', 'South Carolina' = 'SC',
'South Dakota' = 'SD', 'Tennessee' = 'TN', 'Texas' = 'TX', 'Utah' = 'UT',
'Vermont' = 'VT', 'Virginia' = 'VA', 'Washington' = 'WA',
'West Virginia' = 'WV', 'Wisconsin' = 'WI', 'Wyoming' = 'WY'
)

# add abbreviation column
vacc_df <- vacc_df %>%
  mutate(State = state_abbrev[Geography])
write_csv(vacc_df, "vaccination.csv")

```

## Urban-Rural Processing

```

# Download Urban-Rural
url <- "https://www.cdc.gov/nchs/data/data-analysis/NCHSurb-rural-codes.csv"
download.file(url, destfile = "urban_rural.xlsx", mode = "wb")
urban_rural <- read_csv("/Users/yuxinyi/Dartmouth/data wrangling/finalProject/urban_rural.csv")

## Rows: 3160 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr (3): ST_ABBREV, CTYNAME, CBSATITLE
## dbl (8): STFIPS, CTYFIPS, CBSAPOP, CTYPOP, CODE2023, CODE2013, CODE2006, COD...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

#eliminate NA
urban_rural <- urban_rural %>%
  select(ST_ABBREV, CTYPOP, CODE2013, CODE2023) %>%
  filter(!is.na(CTYPOP), !is.na(CODE2013), !is.na(CODE2023))

# Calculate total state population
state_pop <- urban_rural %>%
  group_by(ST_ABBREV) %>%
  summarise(STATEPOP = sum(CTYPOP, na.rm = TRUE))

# Combine the total population and calculate the weight
urban_rural <- urban_rural %>%
  left_join(state_pop, by = "ST_ABBREV") %>%
  mutate(WEIGHT = CTYPOP / STATEPOP,
        INDEX2013 = CODE2013 * WEIGHT,
        INDEX2023 = CODE2023 * WEIGHT)

# Calculation of weighted urban and rural index & changes in 2013 and 2023
urban_result <- urban_rural %>%
  group_by(ST_ABBREV) %>%

```

```

summarise(INDEX2013 = sum(INDEX2013),
           INDEX2023 = sum(INDEX2023)) %>%
mutate(CHANGE = INDEX2023 - INDEX2013) %>%
arrange(desc(CHANGE))

write.csv(urban_result, "state_urban_index_2013_2023.csv", row.names = FALSE)

```

insurance coverage

```

folder <- "/Users/yuxinyi/Dartmouth/data wrangling/finalProject/RawData"
years <- setdiff(2012:2022, 2020) # skip year 2020

# State name to abbreviation
state_abbrev <- c(
  'Alabama' = 'AL', 'Alaska' = 'AK', 'Arizona' = 'AZ', 'Arkansas' = 'AR',
  'California' = 'CA', 'Colorado' = 'CO', 'Connecticut' = 'CT', 'Delaware' = 'DE',
  'District of Columbia' = 'DC', 'Florida' = 'FL', 'Georgia' = 'GA', 'Hawaii' = 'HI',
  'Idaho' = 'ID', 'Illinois' = 'IL', 'Indiana' = 'IN', 'Iowa' = 'IA', 'Kansas' = 'KS',
  'Kentucky' = 'KY', 'Louisiana' = 'LA', 'Maine' = 'ME', 'Maryland' = 'MD',
  'Massachusetts' = 'MA', 'Michigan' = 'MI', 'Minnesota' = 'MN',
  'Mississippi' = 'MS', 'Missouri' = 'MO', 'Montana' = 'MT', 'Nebraska' = 'NE',
  'Nevada' = 'NV', 'New Hampshire' = 'NH', 'New Jersey' = 'NJ',
  'New Mexico' = 'NM', 'New York' = 'NY', 'North Carolina' = 'NC',
  'North Dakota' = 'ND', 'Ohio' = 'OH', 'Oklahoma' = 'OK', 'Oregon' = 'OR',
  'Pennsylvania' = 'PA', 'Rhode Island' = 'RI', 'South Carolina' = 'SC',
  'South Dakota' = 'SD', 'Tennessee' = 'TN', 'Texas' = 'TX', 'Utah' = 'UT',
  'Vermont' = 'VT', 'Virginia' = 'VA', 'Washington' = 'WA',
  'West Virginia' = 'WV', 'Wisconsin' = 'WI', 'Wyoming' = 'WY'
)

all_data <- list()

#Loop to read all year data in one go
for (year in years) {
  file_path <- file.path(folder, paste0("raw_data", year, ".csv"))

  try({
    #Use col_types to explicitly specify the format and skip the Footnotes and Total columns (not used)
    df <- read_csv(
      file_path,
      skip = 2,
      col_types = cols(
        Location = col_character(),
        Employer = col_double(),
        `Non-Group` = col_double(),
        Medicaid = col_double(),
        Medicare = col_double(),
        Military = col_double(),
        Uninsured = col_double(),
        Total = col_skip(),
        Footnotes = col_skip()
      )
    )
  })

```

```

# Keep only the state-level rows and remove the united state row.
df <- df %>%
  filter(Location %in% names(state_abbrev))
#Add year and state abbreviations
df <- df %>%
  mutate(
    Year = year,
    State = state_abbrev[Location],
    Insured = Employer + `Non-Group` + Medicaid + Medicare + Military
  ) %>%
  select(Year, State, Insured, Uninsured)

#summarizing
all_data[[as.character(year)]] <- df
message("Processed ", year, " with ", nrow(df), " rows")

}, silent = TRUE)
}

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2012 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2013 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2014 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2015 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

```

```

## Processed 2016 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2017 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2018 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2019 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2021 with 51 rows

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Processed 2022 with 51 rows

#merge
insurance <- bind_rows(all_data)

write_csv(insurance, "insurance_summary_by_state_2012_2022.csv")

merge

#urban_rural only retains 2023. urban_merge is the version used for merge and is not saved as csv.
urban_merge <- urban_result %>%
  select(ST_ABBREV, INDEX2023) %>%
  rename(State = ST_ABBREV,
         UrbanicityIndex = INDEX2023)
#merge
merged_df <- vacc_df %>%

```

```

mutate(Year = as.numeric(`Survey.Year.Influenza.Season`)) %>%
  left_join(insurance, by = c("State", "Year")) %>%
  left_join(urban_merge, by = "State")
head(merged_df)

##      Vaccine Geography.Type Geography Survey.Year.Influenza.Season
## 1 Influenza          States    Oklahoma             2021
## 2 Influenza          States    Oklahoma             2021
## 3 Influenza          States    Oklahoma             2021
## 4 Influenza          States    Oklahoma             2021
## 5 Tdap               States    Oklahoma             2021
## 6 Tdap               States    Oklahoma             2021
##           Dimension.Type                               Dimension Estimate....
## 1 Race and Ethnicity        White, Non-Hispanic     45.2
## 2 Race and Ethnicity        Black, Non-Hispanic    34.1
## 3 Race and Ethnicity        Hispanic                53.4
## 4 Race and Ethnicity Other or Multiple Races, Non-Hispanic 51.0
## 5 Race and Ethnicity        White, Non-Hispanic    64.9
## 6 Race and Ethnicity        Black, Non-Hispanic    59.6
##   X95..CI.... Sample.Size State Year Insured Uninsured UrbanicityIndex
## 1 38.5 to 52.0         232    OK 2021  933100   201300    2.85059
## 2 28.7 to 39.8         299    OK 2021  933100   201300    2.85059
## 3 48.2 to 58.5         378    OK 2021  933100   201300    2.85059
## 4 45.2 to 56.8         501    OK 2021  933100   201300    2.85059
## 5 57.9 to 71.4         220    OK 2021  933100   201300    2.85059
## 6 53.5 to 65.5         274    OK 2021  933100   201300    2.85059

write_csv(merged_df, "MergedVersion.csv")

```

clean & simplify for regression

```

#clean
merged_clean <- merged_df %>%
  filter(!is.na(Estimate....), !is.na(Uninsured), !is.na(Insured),
         !Estimate.... %in% c("NR*", "N/A", "--", "*")) %>% #ignore "NR*" in Estimate.... due to "NR*"
  mutate(
    Year = as.numeric(Survey.Year.Influenza.Season),
    Coverage = as.numeric(str_replace(Estimate...., "%", "")),
    UninsuredRate = Uninsured / (Insured + Uninsured)
  )

# simplify
merged_summary <- merged_clean %>%
  group_by(State, Year, Vaccine) %>%
  summarise(
    Coverage = mean(Coverage, na.rm = TRUE),
    UninsuredRate = mean(UninsuredRate, na.rm = TRUE),
    UrbanicityIndex = first(UrbanicityIndex)
  ) %>%
  ungroup()

## `summarise()` has grouped output by 'State', 'Year'. You can override using the
## `groups` argument.

```

```

head(merged_summary)

## # A tibble: 6 x 6
##   State  Year Vaccine Coverage UninsuredRate UrbanicityIndex
##   <chr> <dbl> <chr>     <dbl>        <dbl>            <dbl>
## 1 AK      2015 Influenza    56.9       0.168          4.01
## 2 AK      2016 Influenza    59.6       0.178          4.01
## 3 AK      2019 Influenza    63.5       0.119          4.01
## 4 AL      2014 Influenza    41.9       0.165          3.51
## 5 AL      2015 Influenza    45.3       0.139          3.51
## 6 AL      2017 Influenza    46.7       0.133          3.51

```

Influenza visualization&modeling

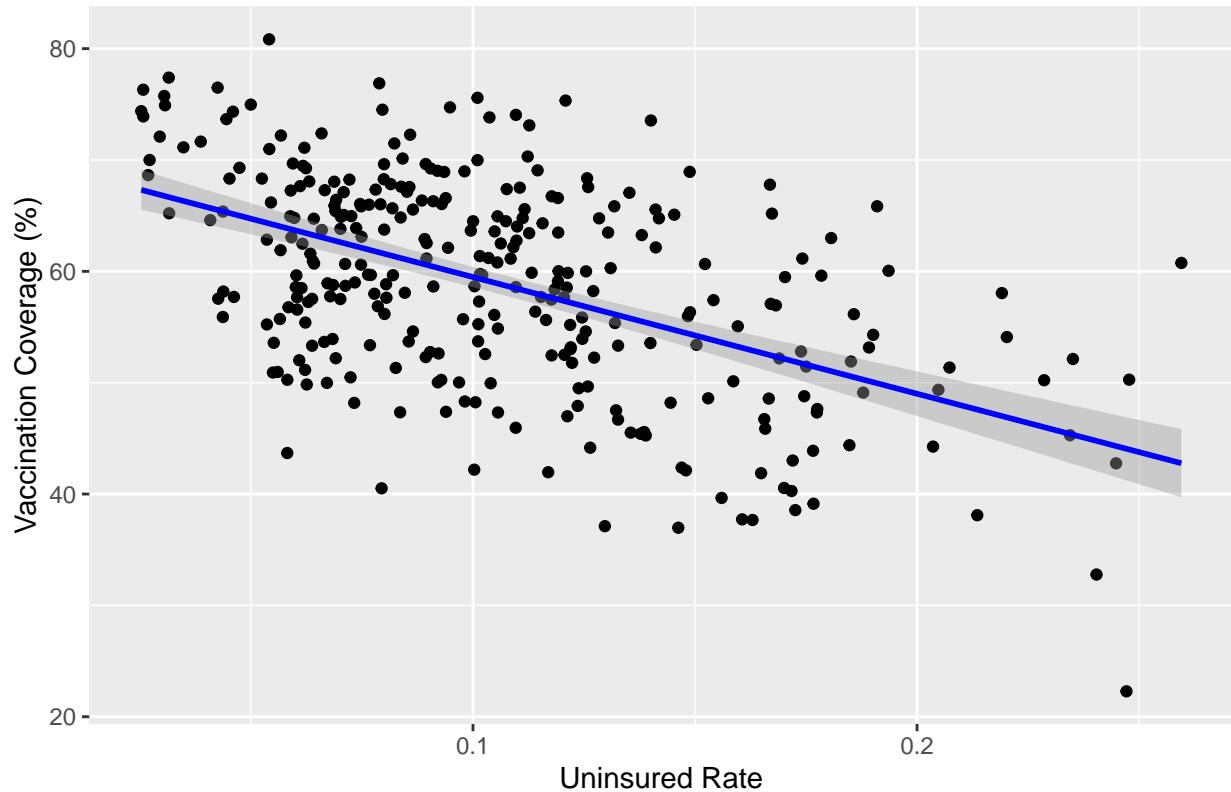
```

merged_summary %>%
  filter(Vaccine == "Influenza") %>%
  ggplot(aes(x = UninsuredRate, y = Coverage)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "Influenza Vaccination vs Uninsured Rate",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  )

## `geom_smooth()` using formula = 'y ~ x'

```

## Influenza Vaccination vs Uninsured Rate



```
model_flu <- lm(Coverage ~ UninsuredRate, data = merged_summary %>% filter(Vaccine == "Influenza"))
summary(model_flu)
```

```
##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = merged_summary %>%
##     filter(Vaccine == "Influenza"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -21.7793  -5.8826   0.9827   5.9490  18.2624 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  69.968     1.123   62.31 <2e-16 ***
## UninsuredRate -104.807    9.658  -10.85 <2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 8.135 on 311 degrees of freedom
## Multiple R-squared:  0.2747, Adjusted R-squared:  0.2723 
## F-statistic: 117.8 on 1 and 311 DF,  p-value: < 2.2e-16
```

Tdap visualization&modeling

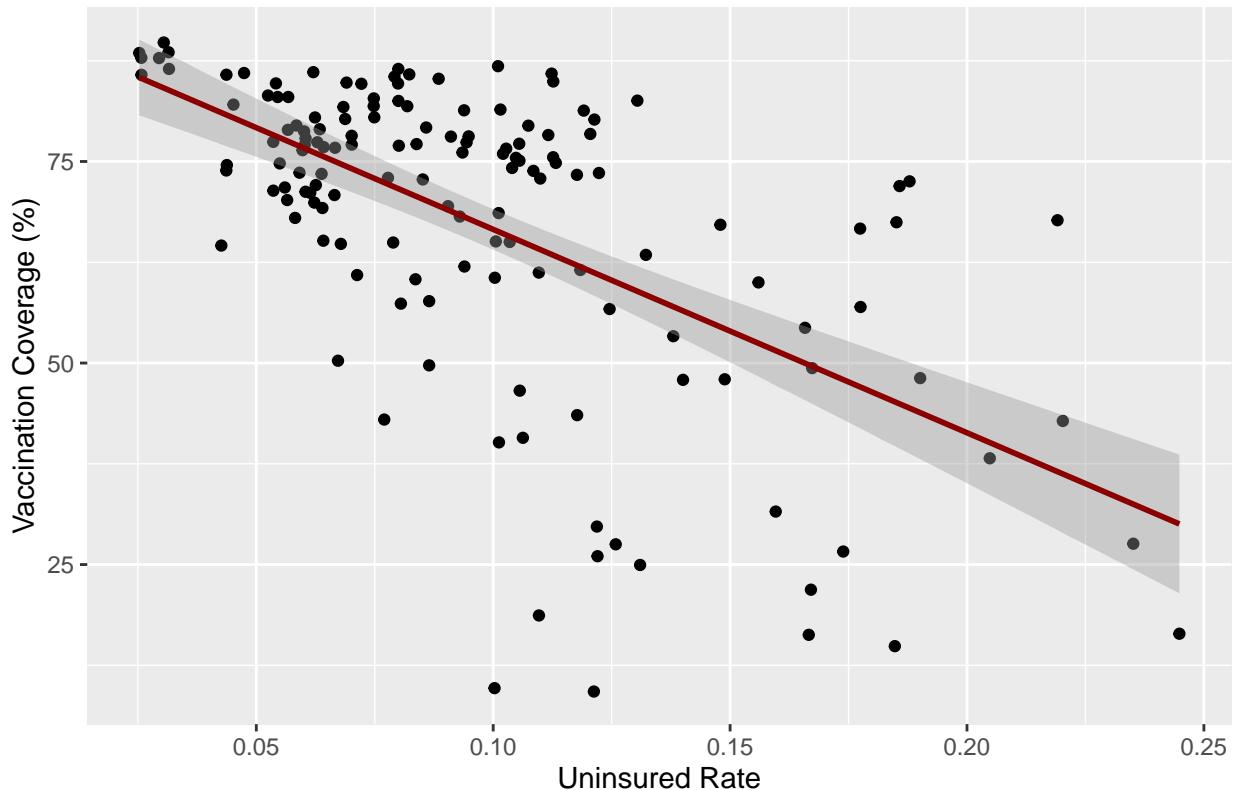
```

merged_summary %>%
  filter(Vaccine == "Tdap") %>%
  ggplot(aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "black") +
  geom_smooth(method = "lm", color = "darkred", se = TRUE) +
  labs(
    title = "Tdap Vaccination vs Uninsured Rate",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  )

```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Tdap Vaccination vs Uninsured Rate



```

tdap_model <- lm(Coverage ~ UninsuredRate, data = merged_summary %>% filter(Vaccine == "Tdap"))
summary(tdap_model)

```

```

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = merged_summary %>%
##     filter(Vaccine == "Tdap"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -56.834  -6.185   2.497  9.916  31.186

```

```

## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)    91.807     3.004  30.560 < 2e-16 ***
## UninsuredRate -252.299    28.136  -8.967 1.47e-15 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 15.31 on 144 degrees of freedom
## Multiple R-squared:  0.3583, Adjusted R-squared:  0.3539 
## F-statistic: 80.41 on 1 and 144 DF,  p-value: 1.467e-15

```

Stratified Analysis by Race for Influenza Vaccination Influenza&race

```

influenza_race <- merged_df %>%
  filter(
    Vaccine == "Influenza",
    Dimension.Type == "Race and Ethnicity",
    !is.na(Estimate....),
    !Estimate.... %in% c("NR*", "N/A", "--", "*")
  ) %>%
  mutate(
    Coverage = as.numeric(str_replace(Estimate...., "%", "")),
    Year = as.numeric(Survey.Year.Influenza.Season),
    UninsuredRate = Uninsured / (Insured + Uninsured)
  )
unique(influenza_race$Dimension)

```

```

## [1] "White, Non-Hispanic"
## [2] "Black, Non-Hispanic"
## [3] "Hispanic"
## [4] "Other or Multiple Races, Non-Hispanic"

```

White, Non-Hispanic

```

influenza_white <- influenza_race %>%
  filter(Dimension == "White, Non-Hispanic")
#plot
ggplot(influenza_white, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "steelblue", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Influenza Coverage vs Uninsured Rate (White, Non-Hispanic)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

```

```
## `geom_smooth()` using formula = 'y ~ x'
```

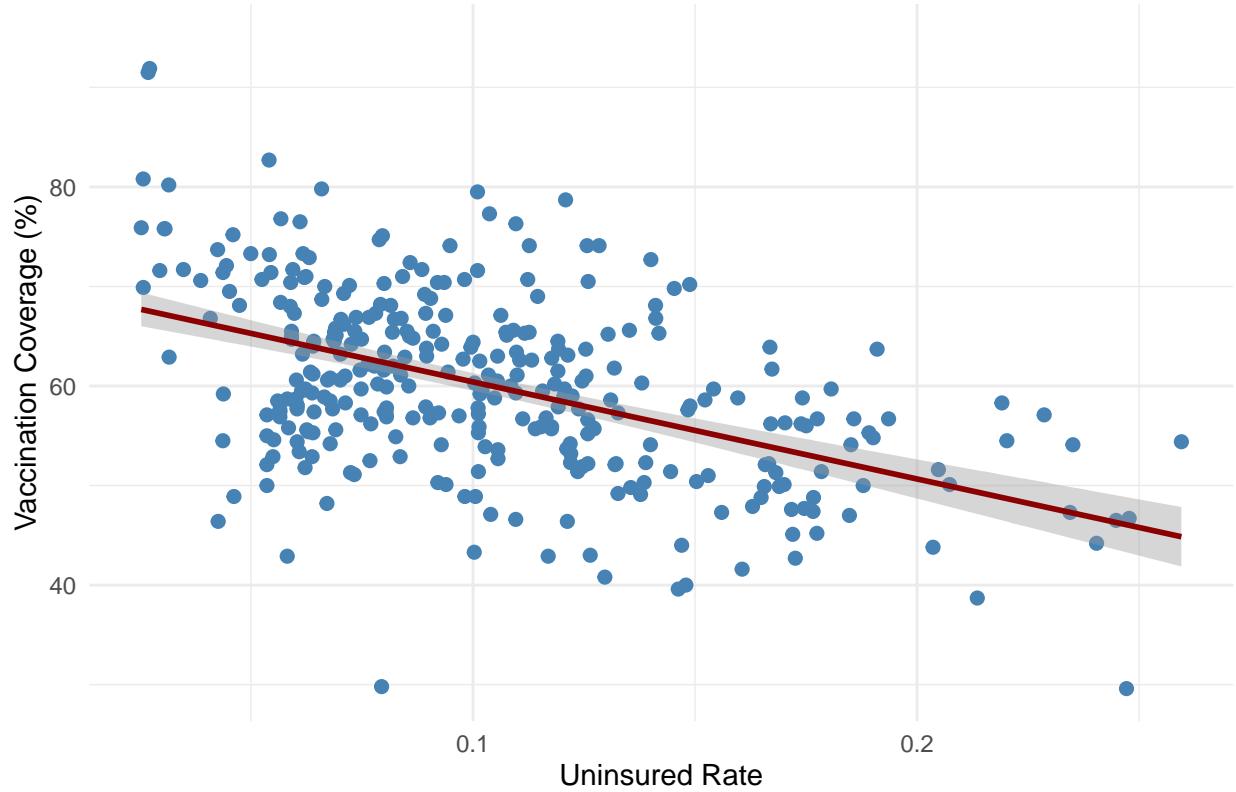
```

## Warning: Removed 57 rows containing non-finite outside the scale range
## ('stat_smooth()').

```

```
## Warning: Removed 57 rows containing missing values or values outside the scale range
## ('geom_point()').
```

### Influenza Coverage vs Uninsured Rate (White, Non-Hispanic)



```
#summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_white))
```

```
##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_white)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -32.618   -5.181    0.360    5.118   24.389 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 70.161     1.071   65.53   <2e-16 ***
## UninsuredRate -97.500     9.363  -10.41   <2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 8.018 on 328 degrees of freedom
## (57 observations deleted due to missingness)
## Multiple R-squared:  0.2485, Adjusted R-squared:  0.2462 
## F-statistic: 108.4 on 1 and 328 DF,  p-value: < 2.2e-16
```

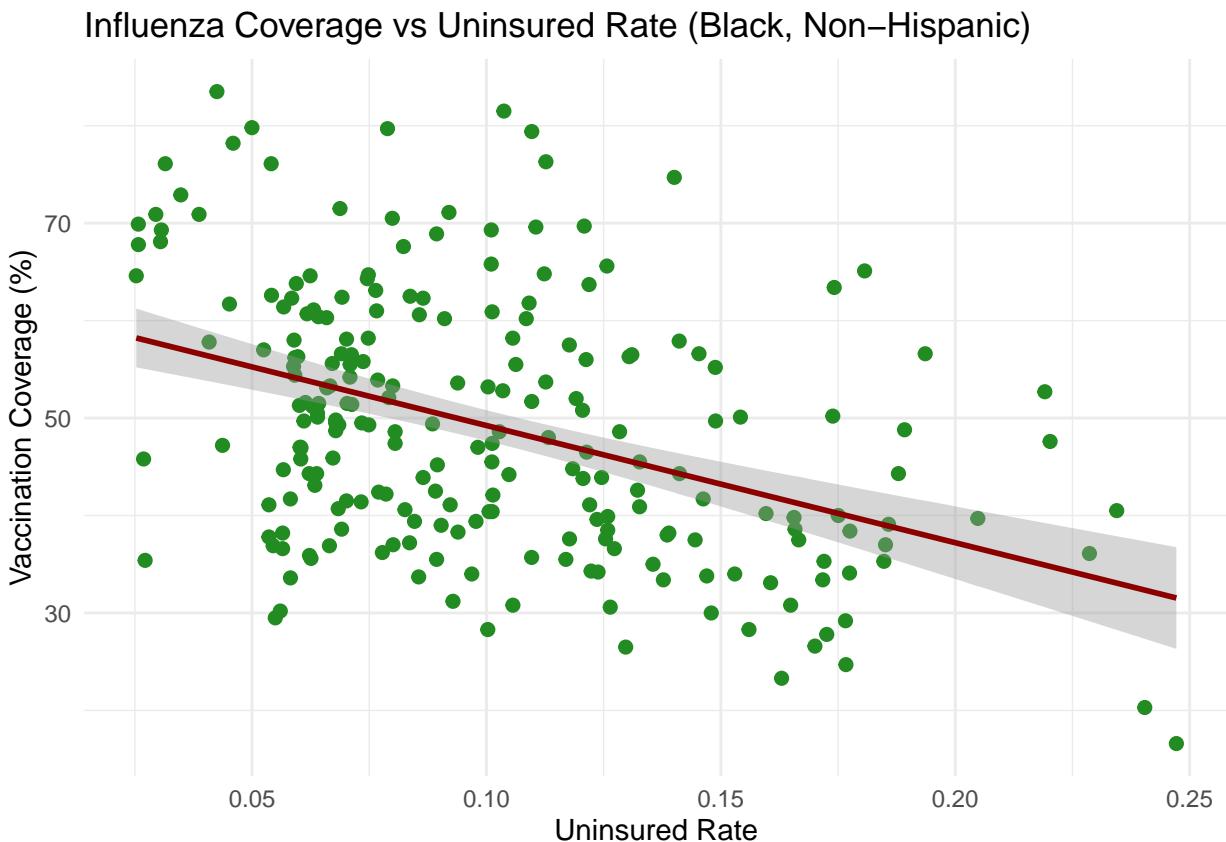
Black, Non-Hispanic

```
influenza_black <- influenza_race %>%
  filter(Dimension == "Black, Non-Hispanic")
#plot
ggplot(influenza_black, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "forestgreen", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Influenza Coverage vs Uninsured Rate (Black, Non-Hispanic)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 31 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 31 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



```

#summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_black))

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_black)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -25.153 -9.355 -1.689  8.946 32.711
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 61.261     1.917 31.949 < 2e-16 ***
## UninsuredRate -120.257    17.275 -6.961 3.59e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.07 on 227 degrees of freedom
##   (31 observations deleted due to missingness)
## Multiple R-squared:  0.1759, Adjusted R-squared:  0.1723
## F-statistic: 48.46 on 1 and 227 DF,  p-value: 3.588e-11

```

Hispanic

```

influenza_hispanic <- influenza_race %>%
  filter(Dimension == "Hispanic")
#plot
ggplot(influenza_hispanic, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "orange", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Influenza Coverage vs Uninsured Rate (Hispanic)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

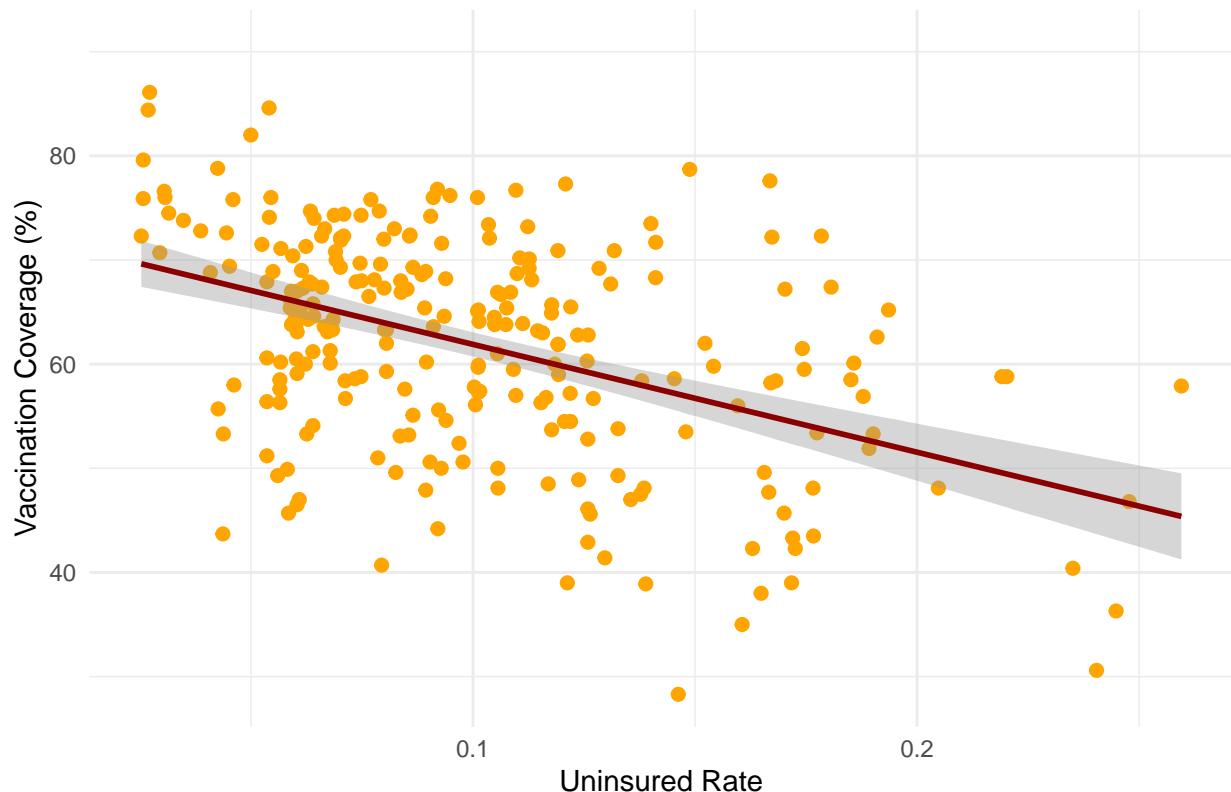
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 34 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 34 rows containing missing values or values outside the scale range
## ('geom_point()').

```

## Influenza Coverage vs Uninsured Rate (Hispanic)



```
#summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_hispanic))
```

```
##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_hispanic)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -28.812  -6.079   1.394   6.157  22.625 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 72.253     1.409  51.283 < 2e-16 ***
## UninsuredRate -103.539    12.686 -8.162 1.56e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.436 on 253 degrees of freedom
## (34 observations deleted due to missingness)
## Multiple R-squared:  0.2084, Adjusted R-squared:  0.2053 
## F-statistic: 66.61 on 1 and 253 DF,  p-value: 1.564e-14
```

Other or Multiple Races, Non-Hispanic

```

influenza_other <- influenza_race %>%
  filter(Dimension == "Other or Multiple Races, Non-Hispanic")
#plot
ggplot(influenza_other, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "purple", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(title = "Influenza Coverage vs Uninsured Rate (Other or Multiple Races, Non-Hispanic)",
       x = "Uninsured Rate",
       y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

```

```

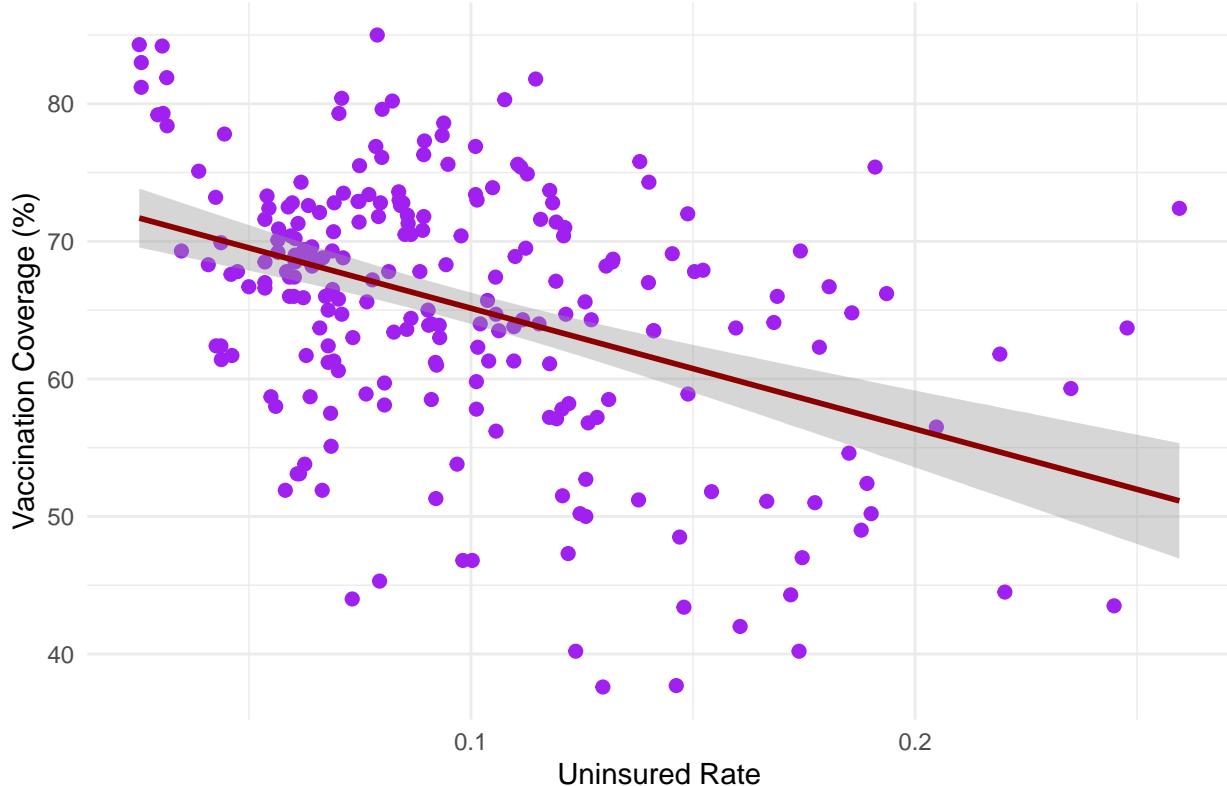
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 40 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 40 rows containing missing values or values outside the scale range
## ('geom_point()').

```

Influenza Coverage vs Uninsured Rate (Other or Multiple Races, Non-Hispanic)



```

#summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_other))

```

```
##
```

```

## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_other)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.9287  -5.3896   0.5326   6.1366  21.2677
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 73.914     1.373  53.825 < 2e-16 ***
## UninsuredRate -87.780    12.709 -6.907 4.89e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.745 on 228 degrees of freedom
##   (40 observations deleted due to missingness)
## Multiple R-squared:  0.173, Adjusted R-squared:  0.1694
## F-statistic:  47.7 on 1 and 228 DF,  p-value: 4.89e-11

```

Merged Influenza Stratified by Race

```

my_colors <- c(
  "White, Non-Hispanic" = "#44d8ee",
  "Black, Non-Hispanic" = "#679436",
  "Hispanic" = "#be0aff",
  "Other or Multiple Races, Non-Hispanic" = "#f46197"
)

ggplot(influenza_race, aes(
  x = UninsuredRate,
  y = Coverage,
  color = Dimension,
  shape = Dimension
)) +
  geom_point(size = 2, alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, fullrange = TRUE) +
  scale_color_manual(values = my_colors) +
  labs(
    title = "Influenza Coverage vs Uninsured Rate by Race/Ethnicity",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)",
    color = "Race/Ethnicity",
    shape = "Race/Ethnicity"
  ) +
  theme_classic() +
  guides(color = guide_legend(override.aes = list(size = 3))) +
  theme(
    legend.text = element_text(size = 5),
    legend.title = element_text(size = 6),
    legend.key.size = unit(0.6, "lines")
  )
}

## `geom_smooth()` using formula = 'y ~ x'

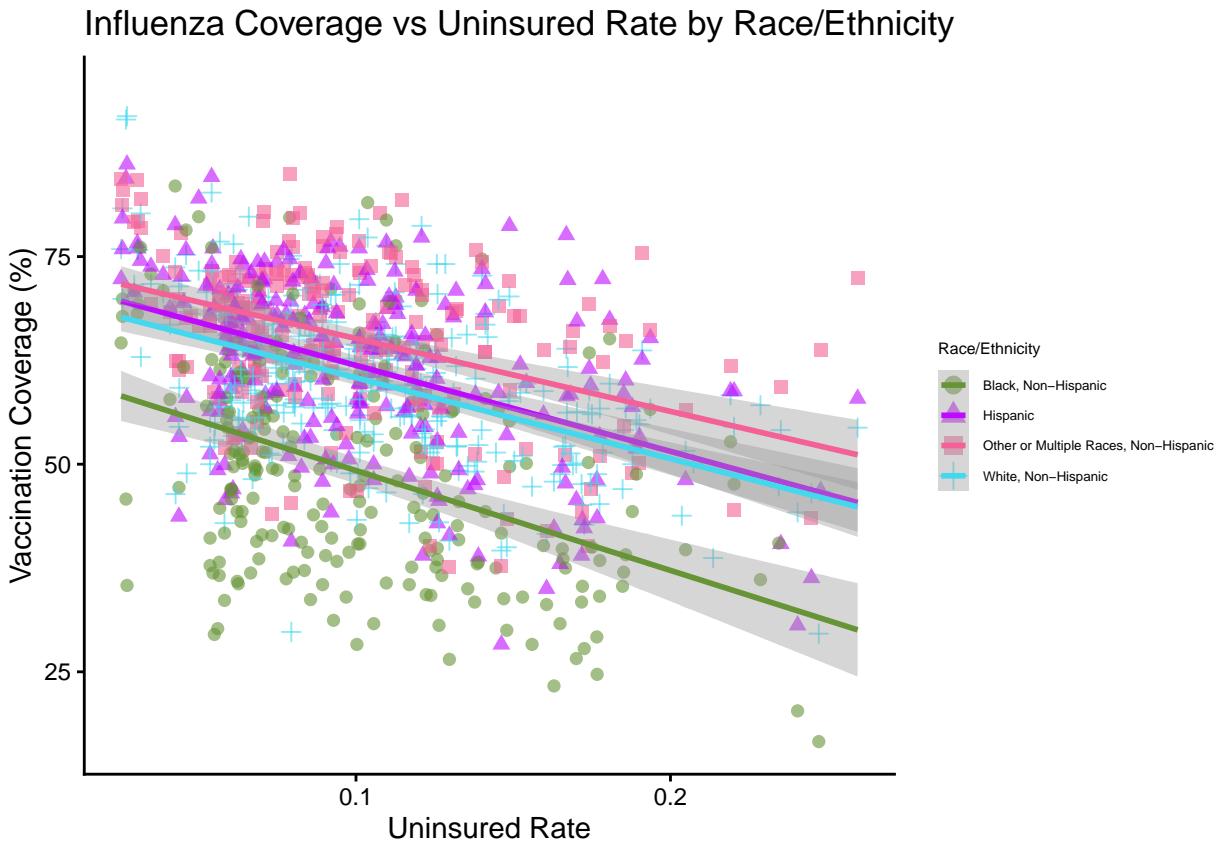
```

```

## Warning: Removed 162 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 162 rows containing missing values or values outside the scale range
## ('geom_point()').

```



Tdap&Race

```

Tdap_race <- merged_df %>%
  filter(
    Vaccine == "Tdap",
    Dimension.Type == "Race and Ethnicity",
    !is.na(Estimate....),
    !Estimate.... %in% c("NR*", "N/A", "--", "*")
  ) %>%
  mutate(
    Coverage = as.numeric(str_replace(Estimate...., "%", "")),
    Year = as.numeric(Survey.Year.Influenza.Season),
    UninsuredRate = Uninsured / (Insured + Uninsured)
  )
unique(Tdap_race$Dimension)

## [1] "White, Non-Hispanic"
## [2] "Black, Non-Hispanic"
## [3] "Hispanic"
## [4] "Other or Multiple Races, Non-Hispanic"

```

White, Non-Hispanic

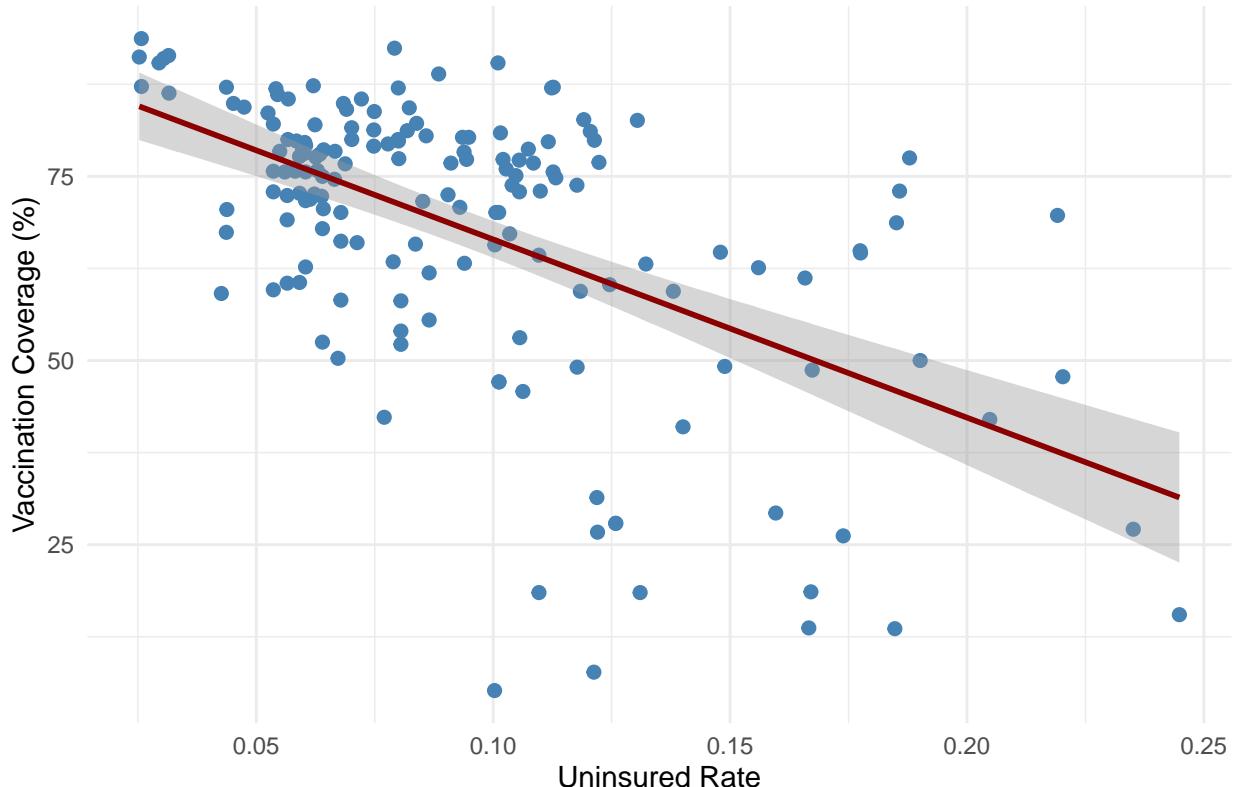
```
Tdap_white <- Tdap_race %>%
  filter(Dimension == "White, Non-Hispanic")
#plot
ggplot(Tdap_white, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "steelblue", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Tdap Coverage vs Uninsured Rate (White, Non-Hispanic)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 26 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 26 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Tdap Coverage vs Uninsured Rate (White, Non-Hispanic)



```

#summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_white))

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_white)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -61.168 -7.106  3.249 10.097 32.318
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  90.62      2.96 30.621 < 2e-16 ***
## UninsuredRate -241.87     28.50 -8.488 1.36e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.94 on 160 degrees of freedom
##   (26 observations deleted due to missingness)
## Multiple R-squared:  0.3105, Adjusted R-squared:  0.3062
## F-statistic: 72.04 on 1 and 160 DF,  p-value: 1.355e-14

```

Black, Non-Hispanic

```

Tdap_black <- Tdap_race %>%
  filter(Dimension == "Black, Non-Hispanic")
#plot
ggplot(Tdap_black, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "forestgreen", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(title = "Tdap Coverage vs Uninsured Rate (Black, Non-Hispanic)",
       x = "Uninsured Rate",
       y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

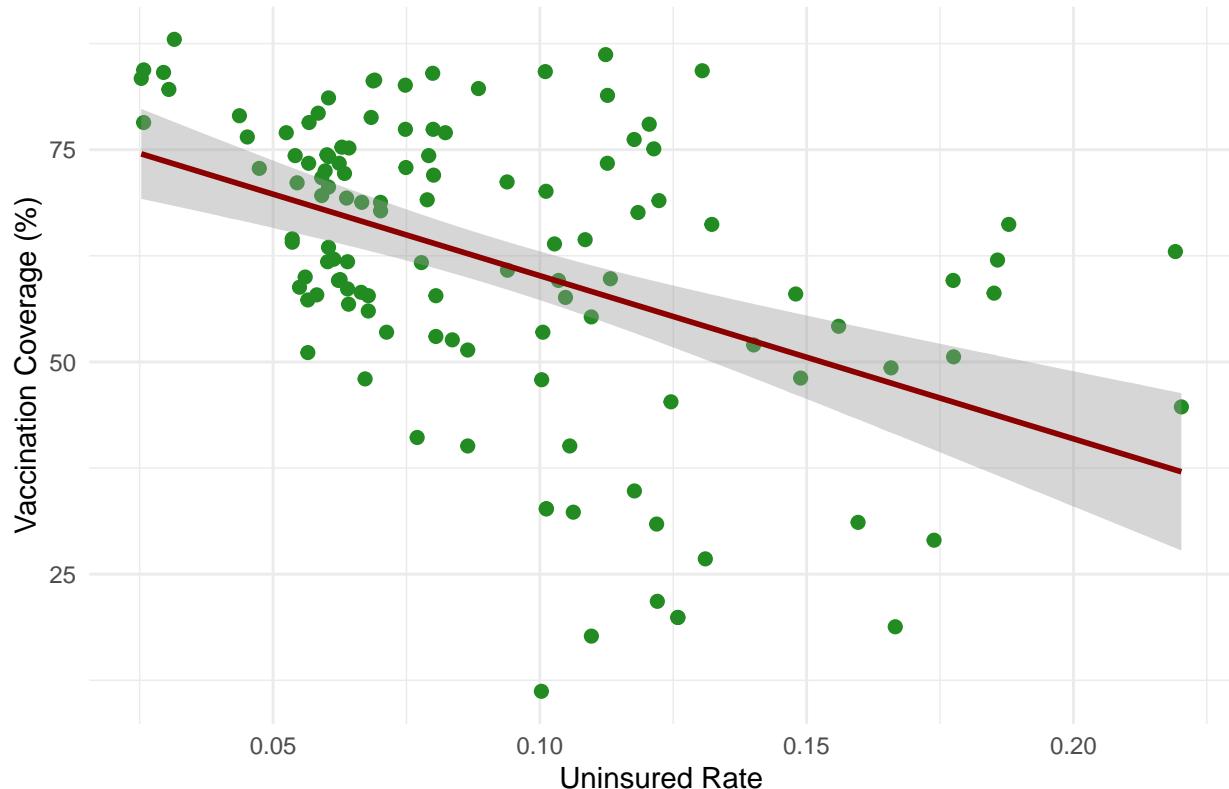
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 15 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 15 rows containing missing values or values outside the scale range
## ('geom_point()').

```

### Tdap Coverage vs Uninsured Rate (Black, Non-Hispanic)



```
#summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_black))
```

```
##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_black)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -48.900  -8.955   2.872  10.150  29.991 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 79.366     3.446  23.029 < 2e-16 ***
## UninsuredRate -192.116    34.516 -5.566 1.66e-07 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 15.54 on 118 degrees of freedom
## (15 observations deleted due to missingness)
## Multiple R-squared:  0.208, Adjusted R-squared:  0.2012 
## F-statistic: 30.98 on 1 and 118 DF,  p-value: 1.658e-07
```

Hispanic

```

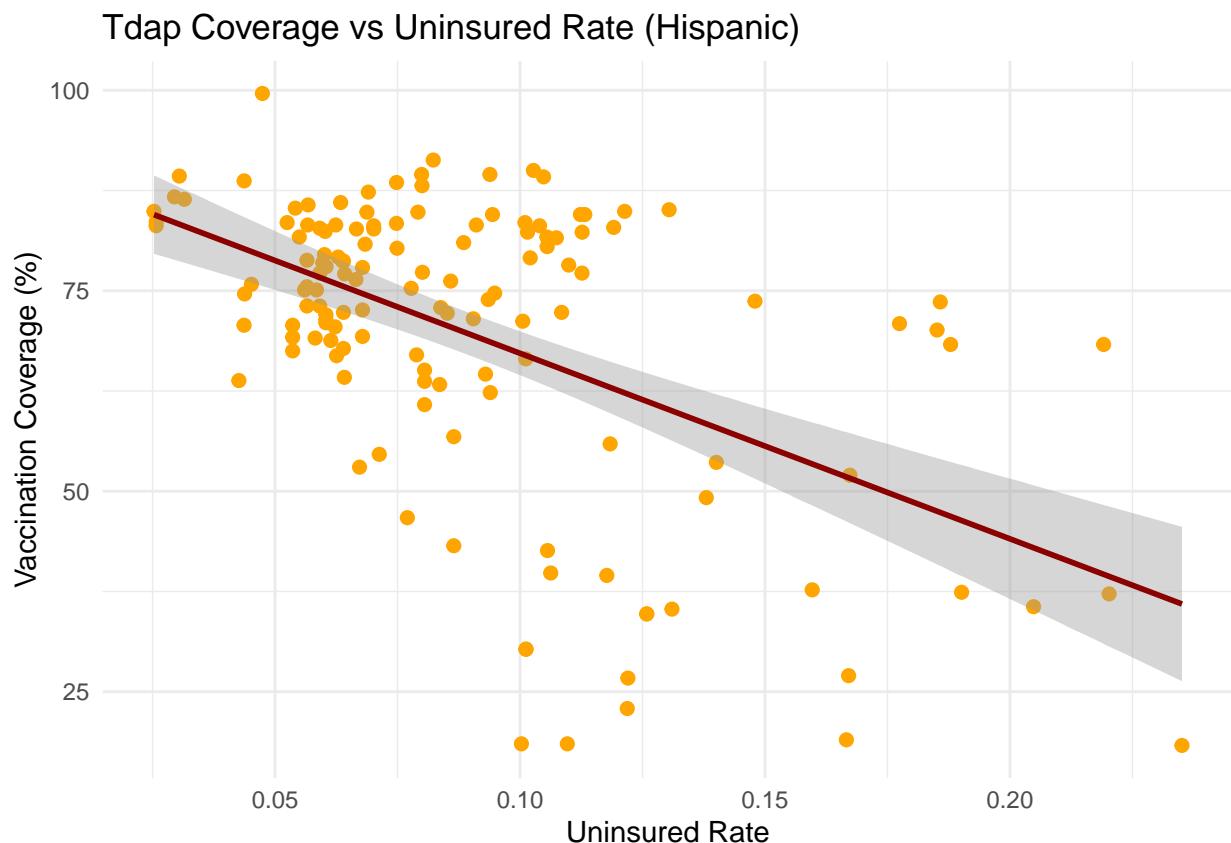
Tdap_hispanic <- Tdap_race %>%
  filter(Dimension == "Hispanic")
#plot
ggplot(Tdap_hispanic, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "orange", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Tdap Coverage vs Uninsured Rate (Hispanic)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 15 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 15 rows containing missing values or values outside the scale range
## ('geom_point()').

```



```

#summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_hispanic))

```

```

## 
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_hispanic)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -48.635  -7.344   1.798  10.369  28.662 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  90.341     3.188  28.337 < 2e-16 ***
## UninsuredRate -231.409    32.207  -7.185 4.01e-11 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 15.75 on 136 degrees of freedom
##   (15 observations deleted due to missingness)
## Multiple R-squared:  0.2752, Adjusted R-squared:  0.2698 
## F-statistic: 51.63 on 1 and 136 DF,  p-value: 4.013e-11

```

Other or Multiple Races, Non-Hispanic

```

Tdap_other <- Tdap_race %>%
  filter(Dimension == "Other or Multiple Races, Non-Hispanic")
#plot
ggplot(Tdap_other, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "purple", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(title = "Tdap Coverage vs Uninsured Rate (Other or Multiple Races, Non-Hispanic)",
       x = "Uninsured Rate",
       y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

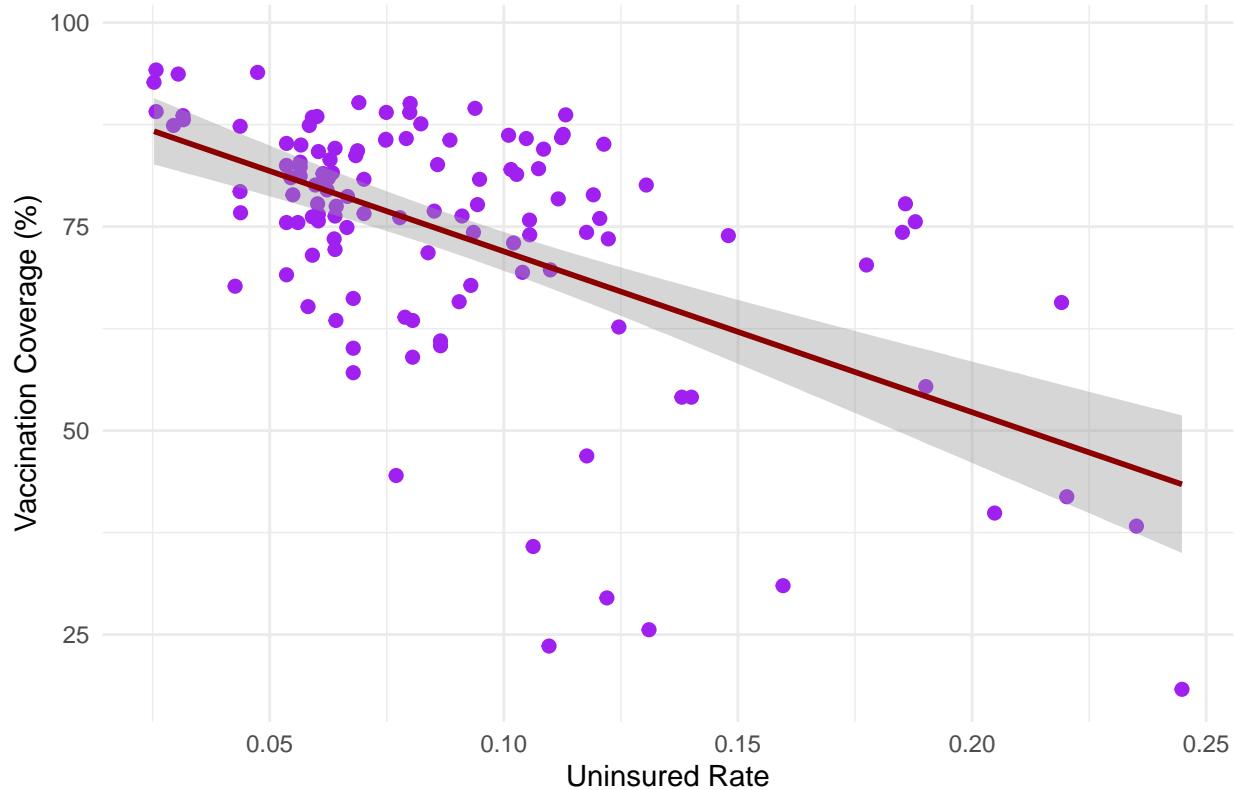
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 22 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 22 rows containing missing values or values outside the scale range
## (`geom_point()`).

```

## Tdap Coverage vs Uninsured Rate (Other or Multiple Races, Non-Hispanic)



```
#summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_other))

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_other)
##
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -46.463 -5.592   1.843   8.593  22.740 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  91.676    2.629  34.865 < 2e-16 ***
## UninsuredRate -197.098   26.311 -7.491 1.27e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.95 on 120 degrees of freedom
## (22 observations deleted due to missingness)
## Multiple R-squared:  0.3186, Adjusted R-squared:  0.313 
## F-statistic: 56.12 on 1 and 120 DF,  p-value: 1.275e-11
```

Merged Tdap Stratified by Race

```

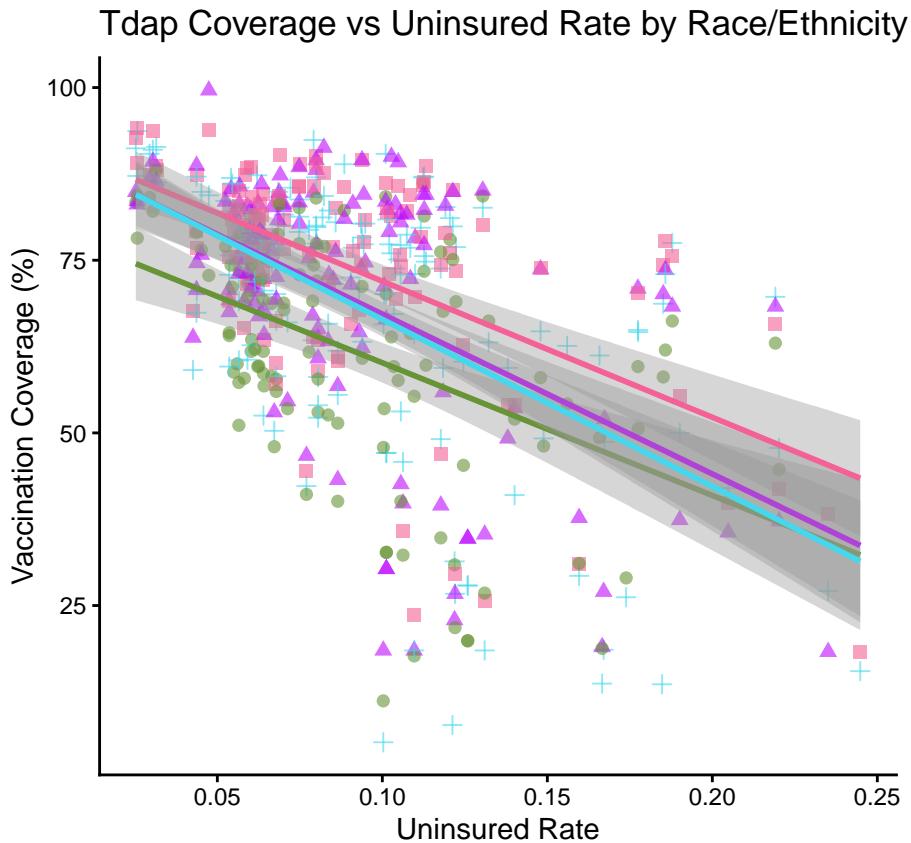
ggplot(Tdap_race, aes(
  x = UninsuredRate,
  y = Coverage,
  color = Dimension,
  shape = Dimension
)) +
  geom_point(size = 2, alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, fullrange = TRUE) +
  scale_color_manual(values = my_colors) +
  labs(
    title = "Tdap Coverage vs Uninsured Rate by Race/Ethnicity",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)",
    color = "Race/Ethnicity",
    shape = "Race/Ethnicity"
  ) +
  theme_classic() +
  guides(color = guide_legend(override.aes = list(size = 3))) +
  theme(
    legend.text = element_text(size = 5),
    legend.title = element_text(size = 6),
    legend.key.size = unit(0.6, "lines")
  )

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 78 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 78 rows containing missing values or values outside the scale range
## ('geom_point()').

```



#### Stratified Analysis by Age Group for Influenza Vaccination Influenza&Age

```
influenza_age <- merged_df %>%
  filter(
    Vaccine == "Influenza",
    Dimension.Type == "Age",
    !is.na(Estimate....),
    !Estimate.... %in% c("NR*", "N/A", "--", "*")
  ) %>%
  mutate(
    Coverage = as.numeric(str_replace(Estimate...., "%", "")),
    Year = as.numeric(Survey.Year.Influenza.Season),
    UninsuredRate = Uninsured / (Insured + Uninsured)
  )
unique(influenza_age$Dimension)
```

```
## [1] ">=18 Years" "18-24 Years" "25-34 Years" ">=35 Years"
```

=18 Years

```
influenza_18 <- influenza_age %>%
  filter(Dimension == ">=18 Years")

# Plot
ggplot(influenza_18, aes(x = UninsuredRate, y = Coverage)) +
```

```

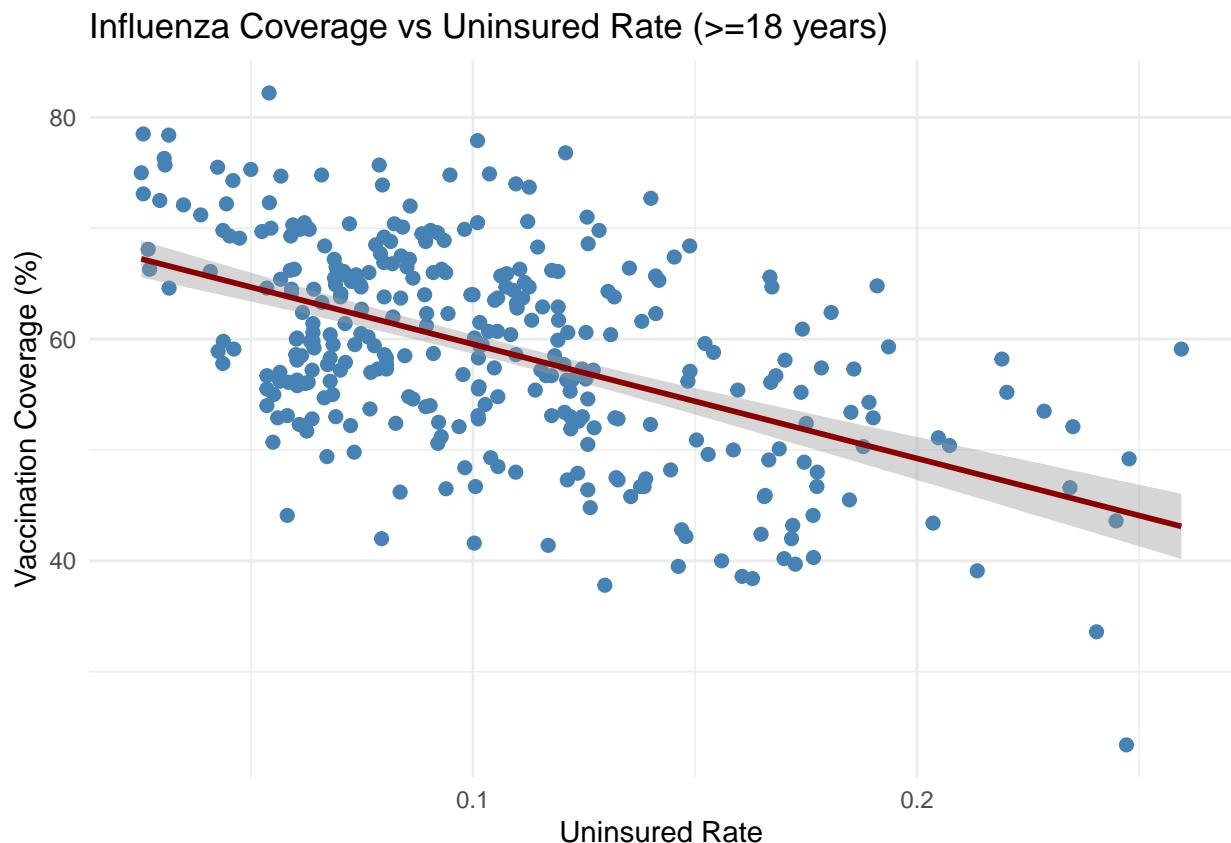
geom_point(color = "steelblue", size = 2) +
geom_smooth(method = "lm", se = TRUE, color = "darkred") +
labs(
  title = "Influenza Coverage vs Uninsured Rate (>=18 years)",
  x = "Uninsured Rate",
  y = "Vaccination Coverage (%)"
) +
theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 57 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 57 rows containing missing values or values outside the scale range
## ('geom_point()').

```



```

# Summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_18))

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_18)

```

```

## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.9809  -6.0249   0.4701   5.8385  19.4104
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  69.836    1.056   66.12 <2e-16 ***
## UninsuredRate -102.988   9.224  -11.17 <2e-16 ***
## ---    
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 7.914 on 329 degrees of freedom
##   (57 observations deleted due to missingness)
## Multiple R-squared:  0.2748, Adjusted R-squared:  0.2726 
## F-statistic: 124.7 on 1 and 329 DF,  p-value: < 2.2e-16

```

18-24 Years

```

influenza_1824 <- influenza_age %>%
  filter(Dimension == "18-24 Years")

# Plot
ggplot(influenza_1824, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "forestgreen", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Influenza Coverage vs Uninsured Rate (18-24 Years)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

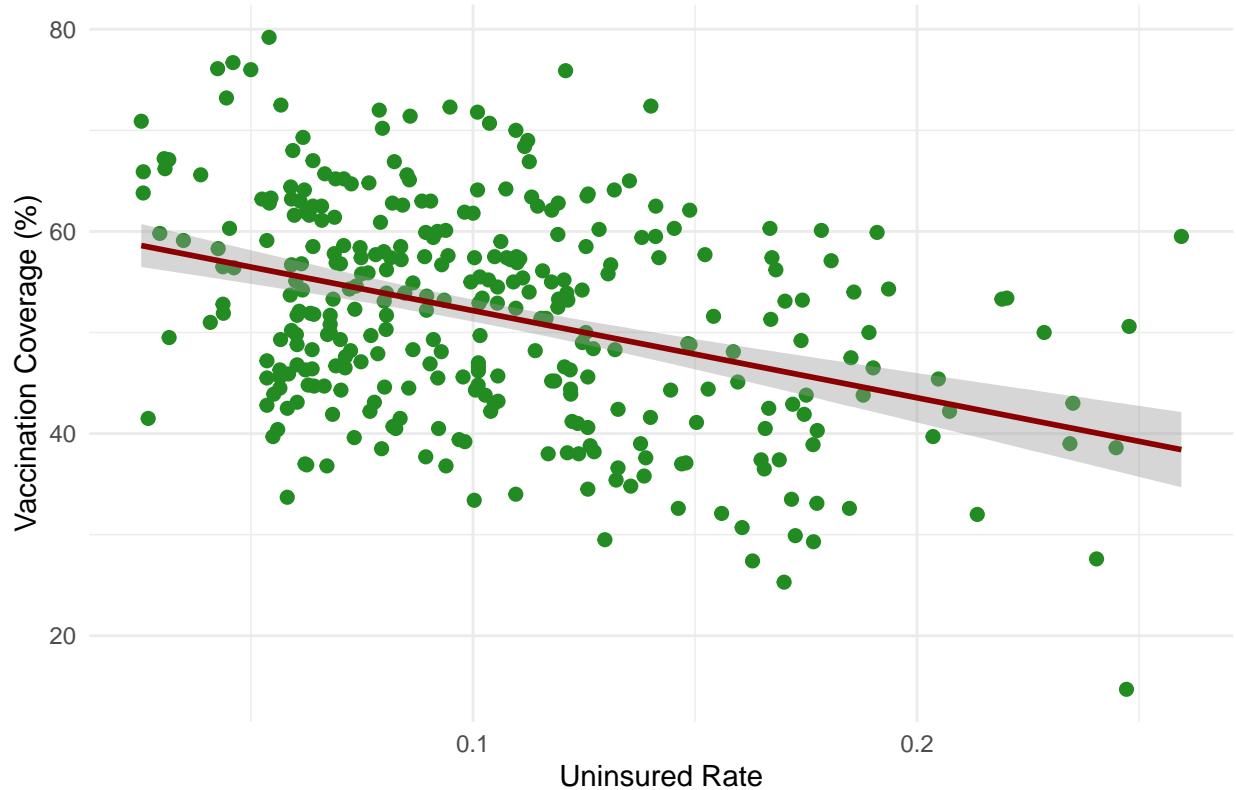
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 56 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 56 rows containing missing values or values outside the scale range
## ('geom_point()').

```

### Influenza Coverage vs Uninsured Rate (18–24 Years)



```
# Summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_1824))
```

```
##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_1824)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -24.7746  -7.8929   0.5474   7.2560  25.5349 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 60.785     1.343   45.246 < 2e-16 ***
## UninsuredRate -86.218    11.692  -7.374 1.38e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.964 on 326 degrees of freedom
## (56 observations deleted due to missingness)
## Multiple R-squared:  0.143, Adjusted R-squared:  0.1403 
## F-statistic: 54.38 on 1 and 326 DF,  p-value: 1.376e-12
```

25-34 Years

```

influenza_2534 <- influenza_age %>%
  filter(Dimension == "25-34 Years")

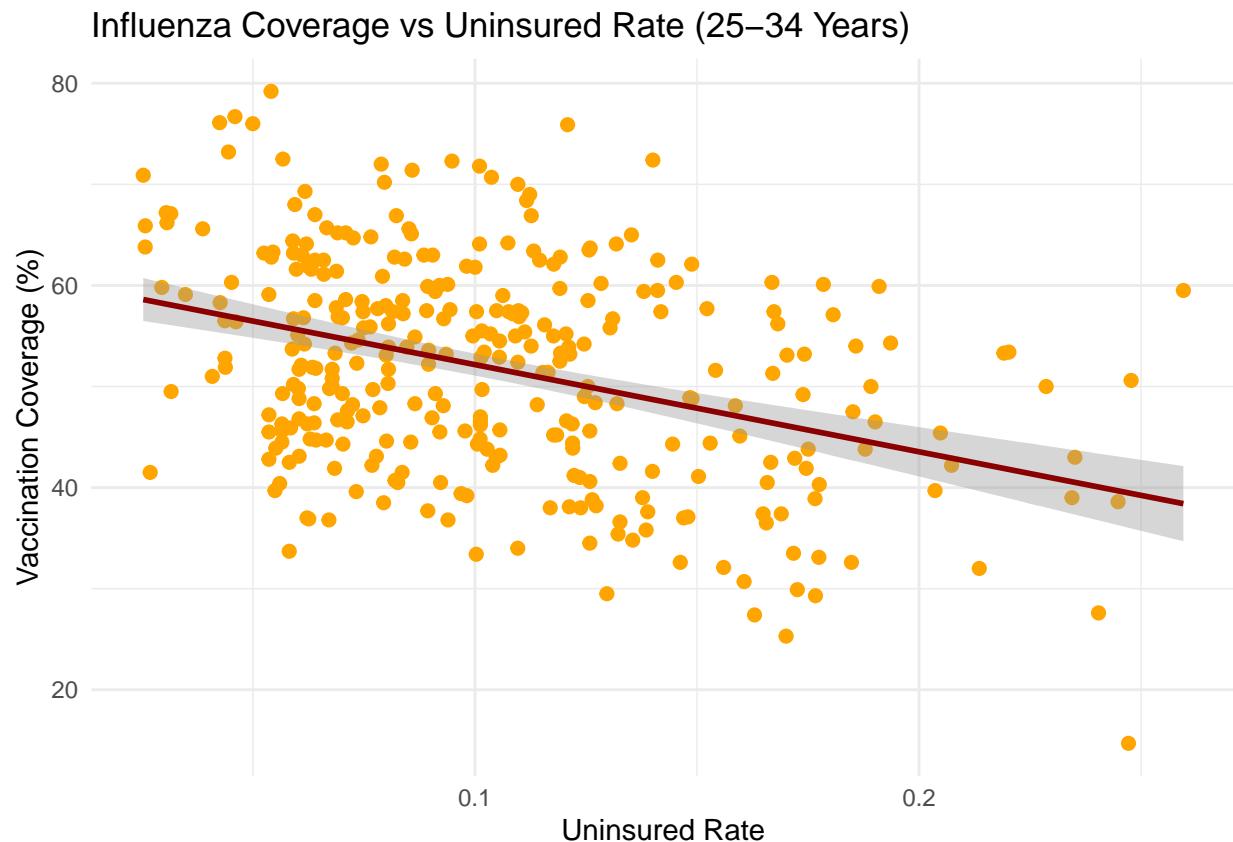
# Plot
ggplot(influenza_1824, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "orange", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Influenza Coverage vs Uninsured Rate (25-34 Years)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 56 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 56 rows containing missing values or values outside the scale range
## ('geom_point()').

```



```

# Summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_2534))

## 
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_2534)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -20.5420  -6.0363   0.5823   5.1607  19.5358 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 70.100     1.029   68.16 <2e-16 ***
## UninsuredRate -91.427    8.982  -10.18 <2e-16 ***  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 7.707 on 329 degrees of freedom
##   (57 observations deleted due to missingness)
## Multiple R-squared:  0.2395, Adjusted R-squared:  0.2372 
## F-statistic: 103.6 on 1 and 329 DF,  p-value: < 2.2e-16
```

=35 Years

```

influenza_35 <- influenza_age %>%
  filter(Dimension == ">=35 Years")

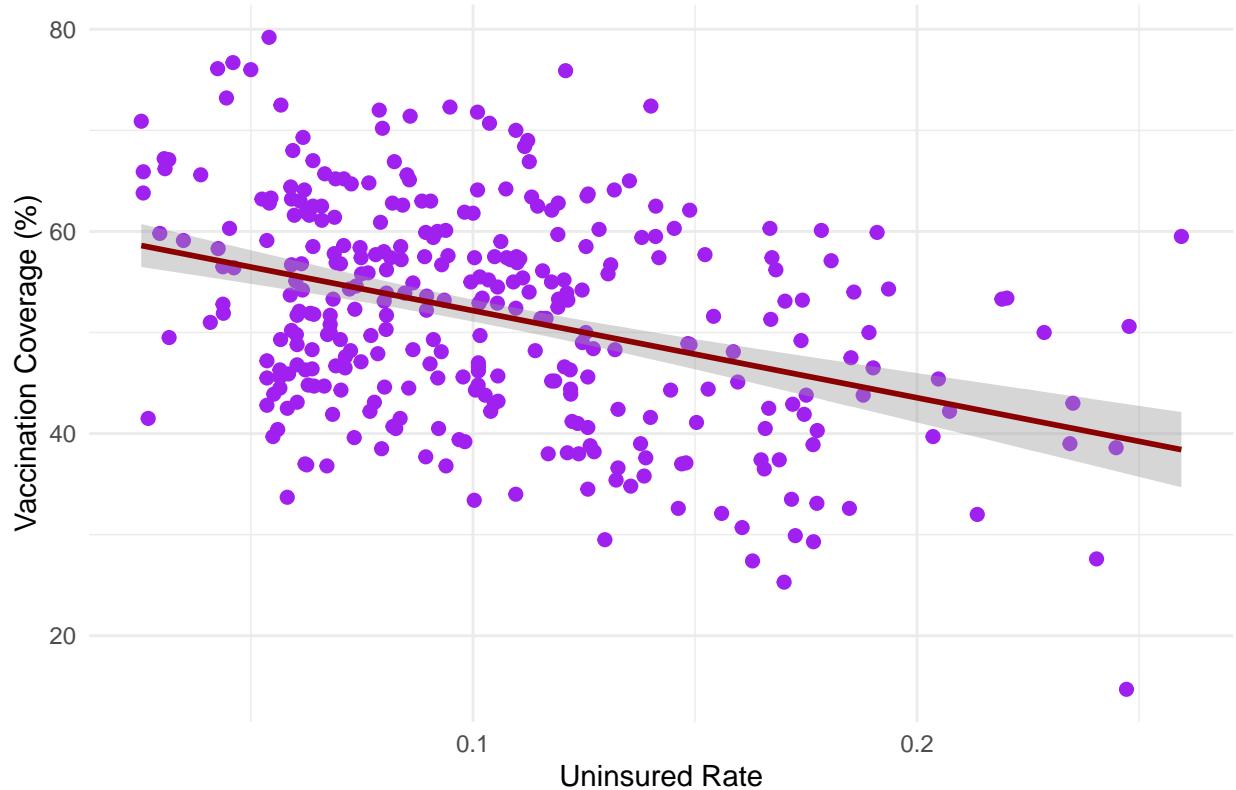
# Plot
ggplot(influenza_1824, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "purple", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Influenza Coverage vs Uninsured Rate (>=35 Years)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 56 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 56 rows containing missing values or values outside the scale range
## ('geom_point()').
```

### Influenza Coverage vs Uninsured Rate (>=35 Years)



```
# Summary
summary(lm(Coverage ~ UninsuredRate, data = influenza_35))
```

```
##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = influenza_35)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -28.2996  -5.7467   0.4052   6.2961  18.5546 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  74.398     1.164   63.93 <2e-16 ***
## UninsuredRate -105.184    10.344  -10.17 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.544 on 318 degrees of freedom
## (56 observations deleted due to missingness)
## Multiple R-squared:  0.2454, Adjusted R-squared:  0.243 
## F-statistic: 103.4 on 1 and 318 DF,  p-value: < 2.2e-16
```

Merged Influenza Stratified by Age Group

```

my_colors_age <- c(
  ">=18 Years" = "#44d8ee",
  "18-24 Years" = "#679436",
  "25-34 Years" = "#be0aff",
  ">=35 Years" = "#f46197"
)
ggplot(influenza_age, aes(
  x = UninsuredRate,
  y = Coverage,
  color = Dimension,
  shape = Dimension
)) +
  geom_point(size = 2, alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, fullrange = TRUE) +
  scale_color_manual(values = my_colors_age) +
  labs(
    title = "Influenza Coverage vs Uninsured Rate by Age Group",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)",
    color = "Age Group",
    shape = "Age Group"
  ) +
  theme_classic()

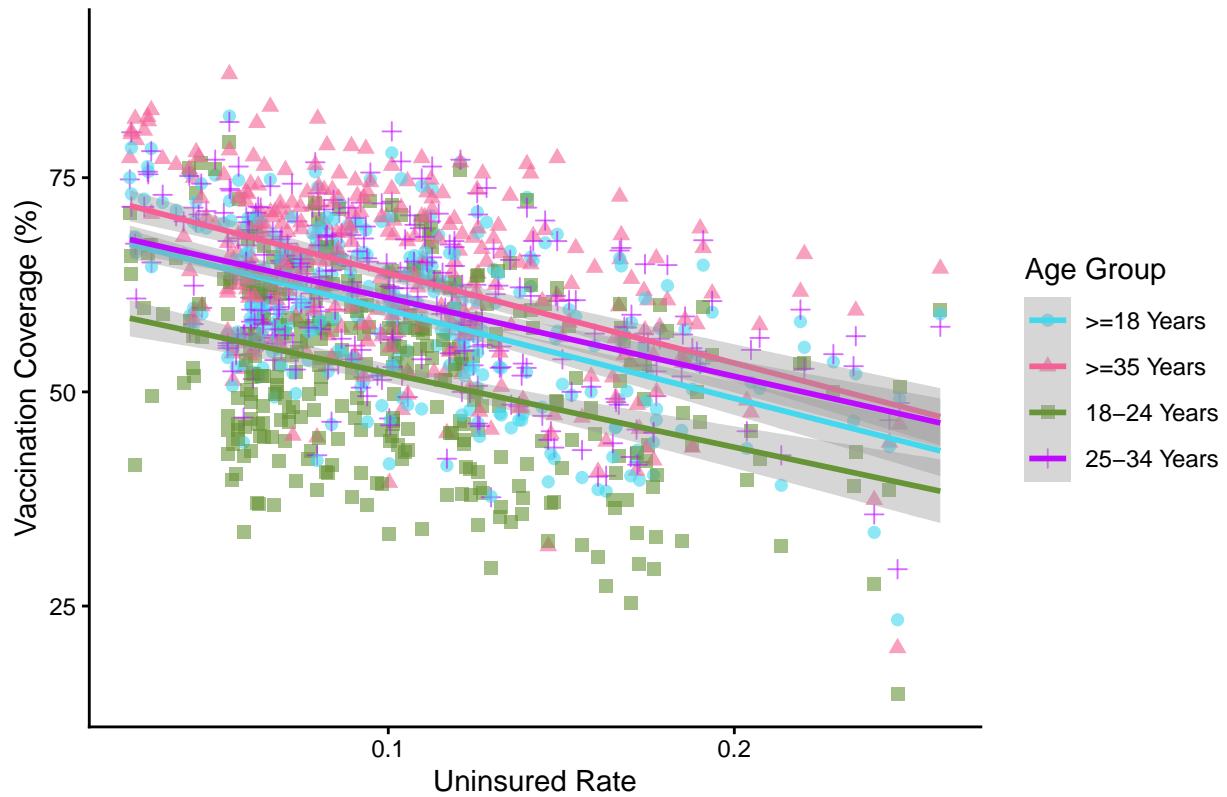
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 226 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 226 rows containing missing values or values outside the scale range
## ('geom_point()').

```

## Influenza Coverage vs Uninsured Rate by Age Group



Stratified Analysis by Age Group for Tdap Vaccination Tdap&Age

```
Tdap_age <- merged_df %>%
  filter(
    Vaccine == "Tdap",
    Dimension.Type == "Age",
    !is.na(Estimate....),
    !Estimate.... %in% c("NR*", "N/A", "--", "*")
  ) %>%
  mutate(
    Coverage = as.numeric(str_replace(Estimate...., "%", "")),
    Year = as.numeric(Survey.Year.Influenza.Season),
    UninsuredRate = Uninsured / (Insured + Uninsured)
  )
unique(Tdap_age$Dimension)
```

```
## [1] ">=18 Years" "18–24 Years" "25–34 Years" ">=35 Years"
```

=18 Years

```
Tdap_18 <- Tdap_age %>%
  filter(Dimension == ">=18 Years")

# Plot
ggplot(Tdap_18, aes(x = UninsuredRate, y = Coverage)) +
```

```

geom_point(color = "steelblue", size = 2) +
geom_smooth(method = "lm", se = TRUE, color = "darkred") +
labs(
  title = "Tdap Coverage vs Uninsured Rate (>=18 years)",
  x = "Uninsured Rate",
  y = "Vaccination Coverage (%)"
) +
theme_minimal()

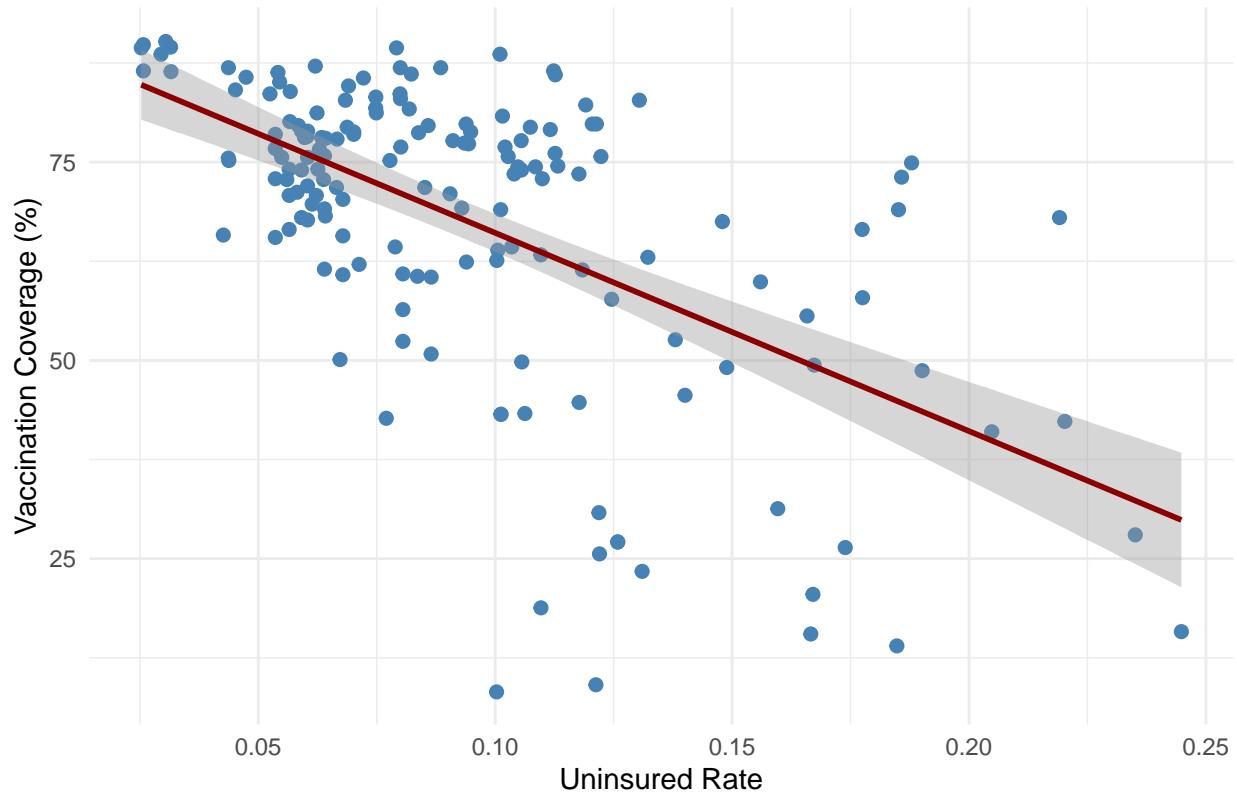
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 26 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 26 rows containing missing values or values outside the scale range
## ('geom_point()').

```

Tdap Coverage vs Uninsured Rate ( $\geq 18$  years)



```

# Summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_18))

```

```

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_18)

```

```

## 
## Residuals:
##    Min     1Q Median     3Q    Max
## -57.804 -6.017  2.762  9.659 31.694
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  91.069    2.842  32.040 < 2e-16 ***
## UninsuredRate -249.934   27.368 -9.132 2.82e-16 ***
## ---    
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 15.3 on 160 degrees of freedom
##   (26 observations deleted due to missingness)
## Multiple R-squared:  0.3426, Adjusted R-squared:  0.3385 
## F-statistic:  83.4 on 1 and 160 DF,  p-value: 2.824e-16

```

18-24 Years

```

Tdap_1824 <- Tdap_age %>%
  filter(Dimension == "18-24 Years")

# Plot
ggplot(Tdap_1824, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "darkgreen", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Tdap Coverage vs Uninsured Rate (18-24 years)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

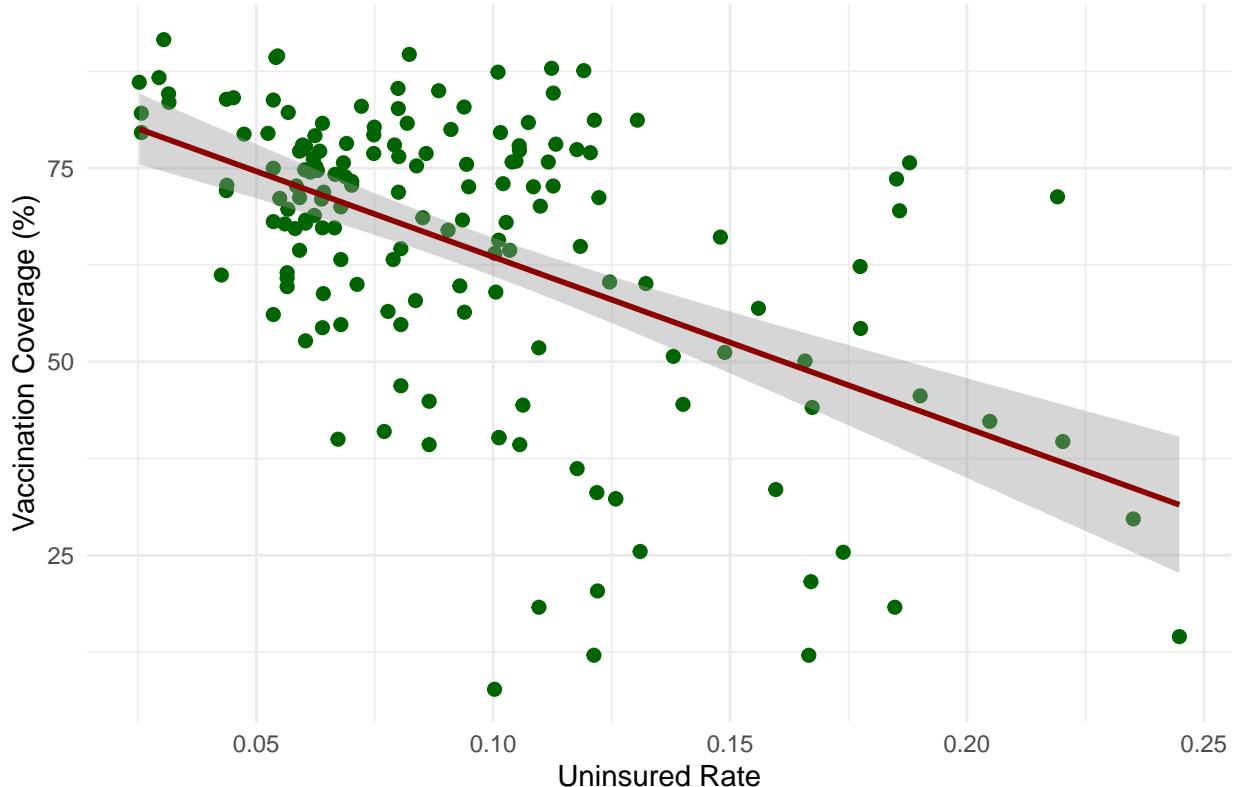
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 26 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 26 rows containing missing values or values outside the scale range
## (`geom_point()`).

```

### Tdap Coverage vs Uninsured Rate (18–24 years)



```
# Summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_1824))

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_1824)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -55.782  -7.998   2.558   9.975  34.085 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  85.650     2.955  28.985 < 2e-16 ***
## UninsuredRate -221.057    28.453 -7.769 8.96e-13 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 15.91 on 160 degrees of freedom
## (26 observations deleted due to missingness)
## Multiple R-squared:  0.2739, Adjusted R-squared:  0.2694 
## F-statistic: 60.36 on 1 and 160 DF,  p-value: 8.96e-13
```

25-34 Years

```

Tdap_2534 <- Tdap_age %>%
  filter(Dimension == "25-34 Years")

# Plot
ggplot(Tdap_2534, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "orange", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Tdap Coverage vs Uninsured Rate (25-34 years)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

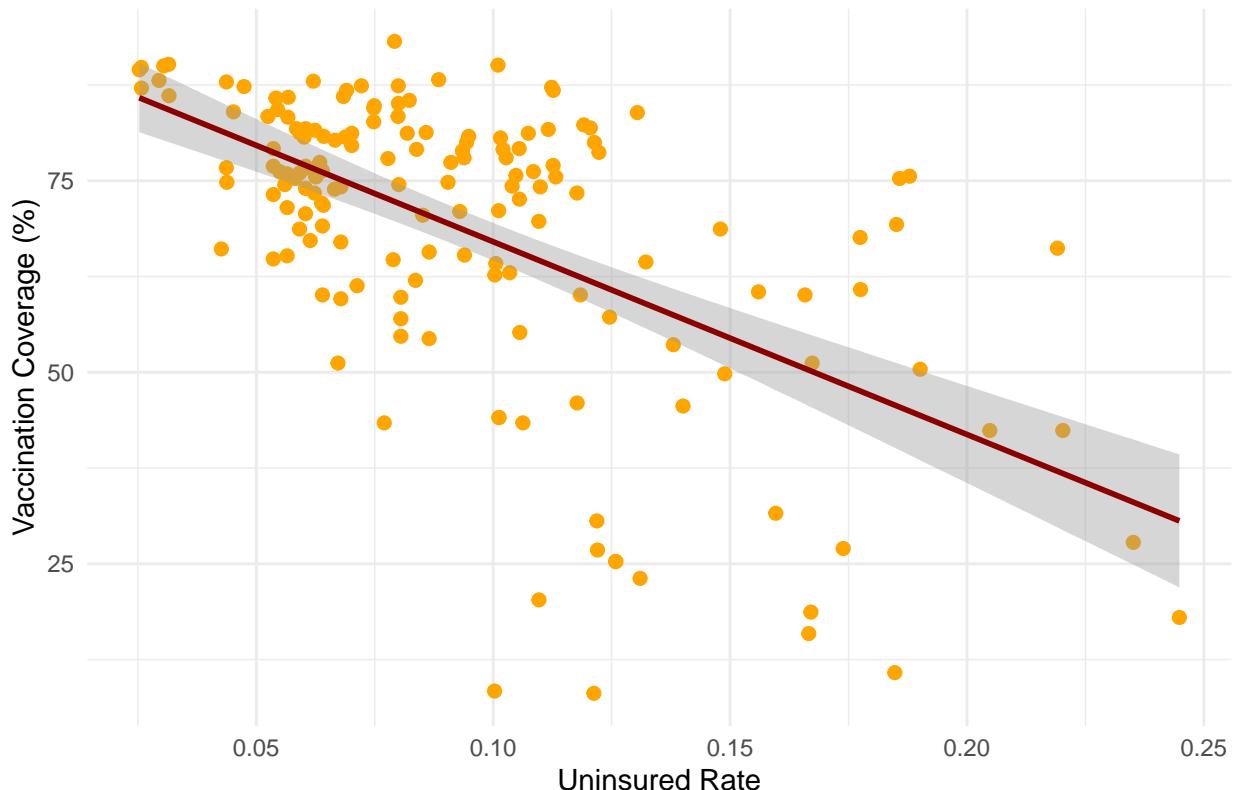
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 26 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 26 rows containing missing values or values outside the scale range
## ('geom_point()').

```

Tdap Coverage vs Uninsured Rate (25–34 years)



```

# Summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_2534))

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_2534)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -58.571 -6.110  3.465 10.136 30.668
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  92.202    2.906  31.729 < 2e-16 ***
## UninsuredRate -251.596   27.980 -8.992 6.6e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.65 on 160 degrees of freedom
##   (26 observations deleted due to missingness)
## Multiple R-squared:  0.3357, Adjusted R-squared:  0.3316
## F-statistic: 80.86 on 1 and 160 DF,  p-value: 6.603e-16

```

=35 Years

```

Tdap_35 <- Tdap_age %>%
  filter(Dimension == ">=35 Years")

# Plot
ggplot(Tdap_35, aes(x = UninsuredRate, y = Coverage)) +
  geom_point(color = "purple", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  labs(
    title = "Tdap Coverage vs Uninsured Rate (>=35 years)",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)"
  ) +
  theme_minimal()

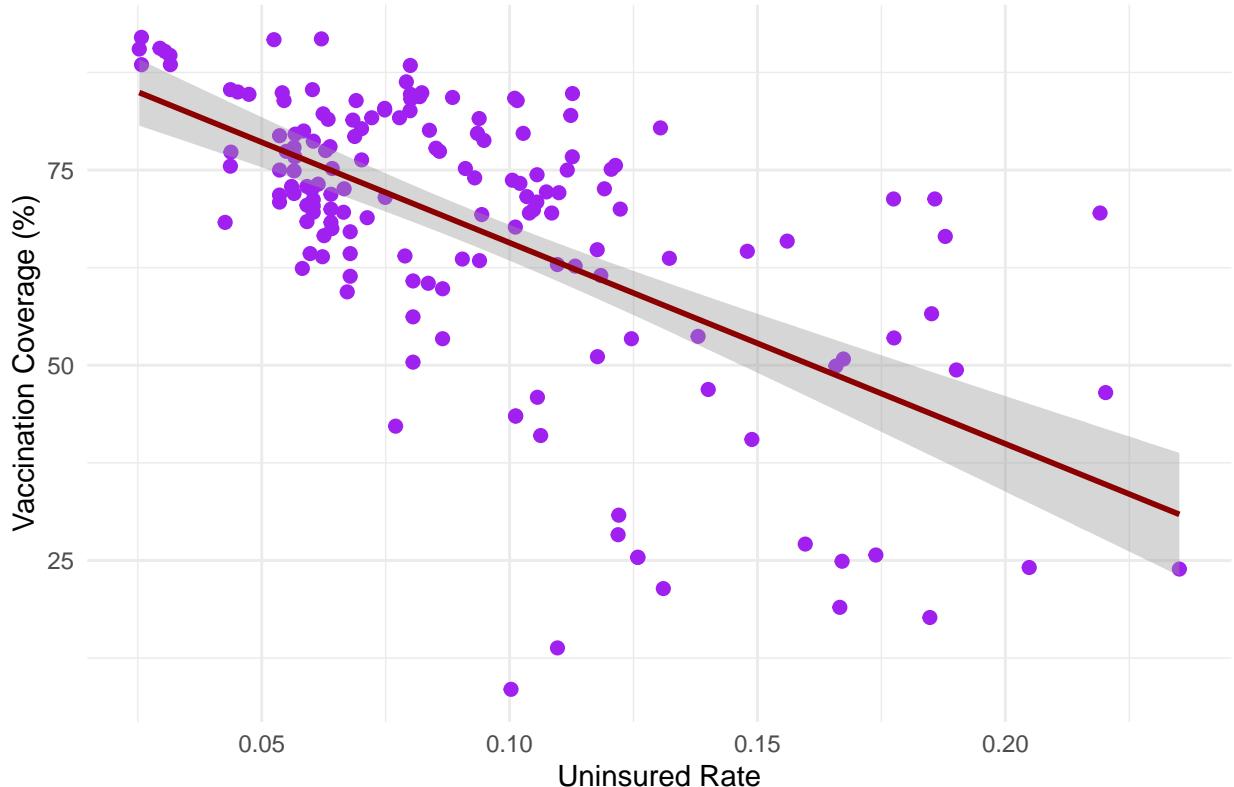
## 'geom_smooth()' using formula = 'y ~ x'

## Warning: Removed 26 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 26 rows containing missing values or values outside the scale range
## ('geom_point()').

```

### Tdap Coverage vs Uninsured Rate ( $\geq 35$ years)



```
# Summary
summary(lm(Coverage ~ UninsuredRate, data = Tdap_35))

##
## Call:
## lm(formula = Coverage ~ UninsuredRate, data = Tdap_35)
##
## Residuals:
##    Min     1Q     Median      3Q     Max 
## -57.117 -6.805   2.771   9.169  34.480 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  91.440    2.748  33.280 <2e-16 ***
## UninsuredRate -257.499   26.869 -9.584 <2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 14.44 on 157 degrees of freedom
## (26 observations deleted due to missingness)
## Multiple R-squared:  0.3691, Adjusted R-squared:  0.3651 
## F-statistic: 91.84 on 1 and 157 DF,  p-value: < 2.2e-16
```

Merged Tdap Stratified by Age Group

```

ggplot(Tdap_age, aes(
  x = UninsuredRate,
  y = Coverage,
  color = Dimension,
  shape = Dimension
)) +
  geom_point(size = 2, alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, fullrange = TRUE) +
  scale_color_manual(values = my_colors_age) +
  labs(
    title = "Tdap Coverage vs Uninsured Rate by Age Group",
    x = "Uninsured Rate",
    y = "Vaccination Coverage (%)",
    color = "Age Group",
    shape = "Age Group"
  ) +
  theme_classic()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 104 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 104 rows containing missing values or values outside the scale range
## (`geom_point()`).

```

