Beyond the Bubble: Understanding American TV News Viewership Patterns

Nolan Windham

12/11/24

Introduction

Television news remains a crucial source of information for many Americans, yet viewing patterns have become increasingly fragmented and potentially polarized. This analysis examines TV news consumption patterns using the Cooperative Election Study (CCES) data from 2020 and 2022, focusing on three key questions:

1. How do ideological preferences relate to news network choice, and to what extent do viewers consume news from multiple sources?
2. What is the age distribution of viewers across different networks, and how might this impact the future of TV news?
3. How closely do viewers' policy preferences align with their chosen news sources, and what does this suggest about media's role in political polarization?

This analysis reveals several important findings. First, while ideological sorting is evident (e.g., Fox News viewers skewing conservative, MSNBC viewers skewing liberal), most networks maintain a significant moderate audience base. Second, TV news viewership skews older across all networks, with median viewer ages ranging from 52-59 years. Finally, it finds strong correlations between viewers' policy preferences and their network choices, particularly on issues like immigration and climate change, though the direction of causality remains unclear.

Data Source

This analysis primarily uses data from the Cooperative Election Study (CCES), a large-scale academic survey project coordinated by Harvard University. The CCES employs an online survey methodology, collecting responses from a nationally stratified sample of approximately 60,000 American adults each year. This analysis combines three key datasets:

1. The Cumulative CCES (2006-2023), which contains 641,955 respondents and harmonizes common variables across survey years. This dataset is available through Harvard Dataverse (doi:10.7910/DVN/II2DB6 (https://doi.org/10.7910/DVN/II2DB6)).

2. Individual year CCES surveys for 2020 and 2022, which provide detailed media consumption variables not included in the cumulative file.

3. The Cumulative CES Policy Preferences dataset (doi:10.7910/DVN/OSXDQO (https://doi.org/10.7910/DVN/OSXDQO)), which offers standardized measures of policy positions.

The CCES employs a two-stage survey design. First, respondents complete a pre-election survey in October, collecting demographic information, political attitudes, and media consumption habits. A post-election wave in November captures voting behavior and additional political opinions. YouGov conducts the survey using matched random sampling to achieve a representative sample of the U.S. adult population.

Data's Strengths: - Large sample size (~60,000 per year) - Consistent core questions across years - Detailed media consumption measures - Comprehensive demographic information

Data's Limitations: - Self-reported media consumption - Online-only survey methodology - Some variables not available in all years - Potential response bias in survey participation

Ethics Reflection

1. **Survey Representation and Access**
    - The CCES's online-only methodology may systematically exclude certain populations, particularly older Americans or those without reliable internet access
    - While YouGov employs matching techniques to achieve representative samples, the initial pool of respondents is limited to those who can and choose to participate in online surveys
    - This could lead to underrepresentation of certain viewpoints and experiences, particularly from rural or low-income communities
2. **Self-Reporting Bias**
    - Media consumption habits are self-reported, which may be subject to social desirability bias
    - Respondents might underreport consumption of politically controversial sources
    - These biases could systematically affect our understanding of media consumption patterns
3. **Privacy and Data Usage**
    - While the CCES data is anonymized, it contains detailed demographic and political information that could potentially be used to identify individuals if combined with other data sources
    - The public availability of this data, while valuable for research, requires some consideration of how findings might be used or misused
    - There is an ethical obligation to present findings in ways that don't stigmatize particular groups or encourage harmful stereotypes. I do my best to avoid this in my analysis, but implicit biases may still be present.
4. **Broader Societal Implications**
    - Research on media consumption patterns could be used to further segment and target audiences, potentially contributing to increased political polarization
    - Studies of partisan media consumption could reinforce existing stereotypes about different political groups

These ethical considerations inform my analysis. First, I am careful to acknowledge the limitations of my data and avoid overgeneralizing my findings. Second, I present results in ways that highlight the complexity of media consumption patterns rather than reducing them to simple partisan narratives. Third, I consider how my findings might be used by others to attempt to frame them in ways that promote understanding rather than division.

Data Import / Cleaning and Tidying

Import packages:

```
library(tidyverse)
library(haven)
```

Import data:

```
# Read in the cumulative CCES data
cc <- read_dta("cumulative_2006-2023.dta")
# Read 2022 CCES data
cc22 <- read_dta("CCES22_Common_OUTPUT_vv_topost.dta")
# Read 2020 CCES data
cc20 <- read_dta("CCES20.dta")
# Read in policy preferences data
policy_prefs <- read_dta("cumulative_ces_policy_preferences.dta")
```

Clean data:

```
# Subset media variables
# NOTE: caseid is the ID variable in the 2022 CCES data
cc22_media <- cc22 |>
  select(caseid, matches("CC22_300"))

# Subset media variables for 2020 data
# NOTE: caseid is the ID variable in the 2020 CCES data
cc20_media <- cc20 |>
  select(caseid, matches("CC20_300"))
```

```
# Helper function to standardize media variables across years
standardize_media_vars <- function(data, year) {
  # This function takes survey data and a year, then:
  # 1. Selects only the case ID and media consumption variables
  # 2. Removes the year prefix from variable names to make them consistent
  # This allows us to combine data across years more easily
  # Select relevant columns and rename them
  prefix <- paste0("CC", year, "_300")

  data |>
    select(caseid, matches(prefix)) |>
    # Rename columns to remove year prefix
    rename_with(~str_replace(., prefix, "300"), matches(prefix))
}

# Process each year's data
cc22_media <- standardize_media_vars(cc22, "22")
cc20_media <- standardize_media_vars(cc20, "20")
```

```r
# Function to merge media data with cumulative data for a specific year
merge_media_data <- function(media_data, year) {
  cc |>
    filter(year == year) |>
    mutate(case_id = as.numeric(case_id)) |>
    left_join(media_data, by = c("case_id" = "caseid"))
}

# Merge data for each year
cc22_merged <- merge_media_data(cc22_media, 2022)
cc20_merged <- merge_media_data(cc20_media, 2020)
```

```r
# Combine 2020 and 2022 data with only necessary variables
cc_media <- bind_rows(
  # 2020 data
  cc20_merged |>
    filter(year == 2020) |>
    select(year, case_id, ideo5, age, matches("300b_[1-8]")),

  # 2022 data
  cc22_merged |>
    filter(year == 2022) |>
    select(year, case_id, ideo5, age, matches("300b_[1-8]"))
)
```

```r
clean_network_names <- function(network_col) {
  case_when(
    network_col == "300b_1" ~ "ABC",
    network_col == "300b_2" ~ "CBS",
    network_col == "300b_3" ~ "NBC",
    network_col == "300b_4" ~ "CNN",
    network_col == "300b_5" ~ "Fox News",
    network_col == "300b_6" ~ "MSNBC",
    network_col == "300b_7" ~ "PBS",
    network_col == "300b_8" ~ "Other",
  )
}
```

```r
# Merge policy preferences with media data
cc_media_policy <- cc_media |>
  filter(year == 2020) |>  # Focus on 2020 data
  # Convert case_id to character for matching
  mutate(case_id = as.character(case_id)) |>
  left_join(
    policy_prefs |>
      select(case_id, enviro_carbon, immig_border, guns_assaultban),
    by = "case_id"
  )
```

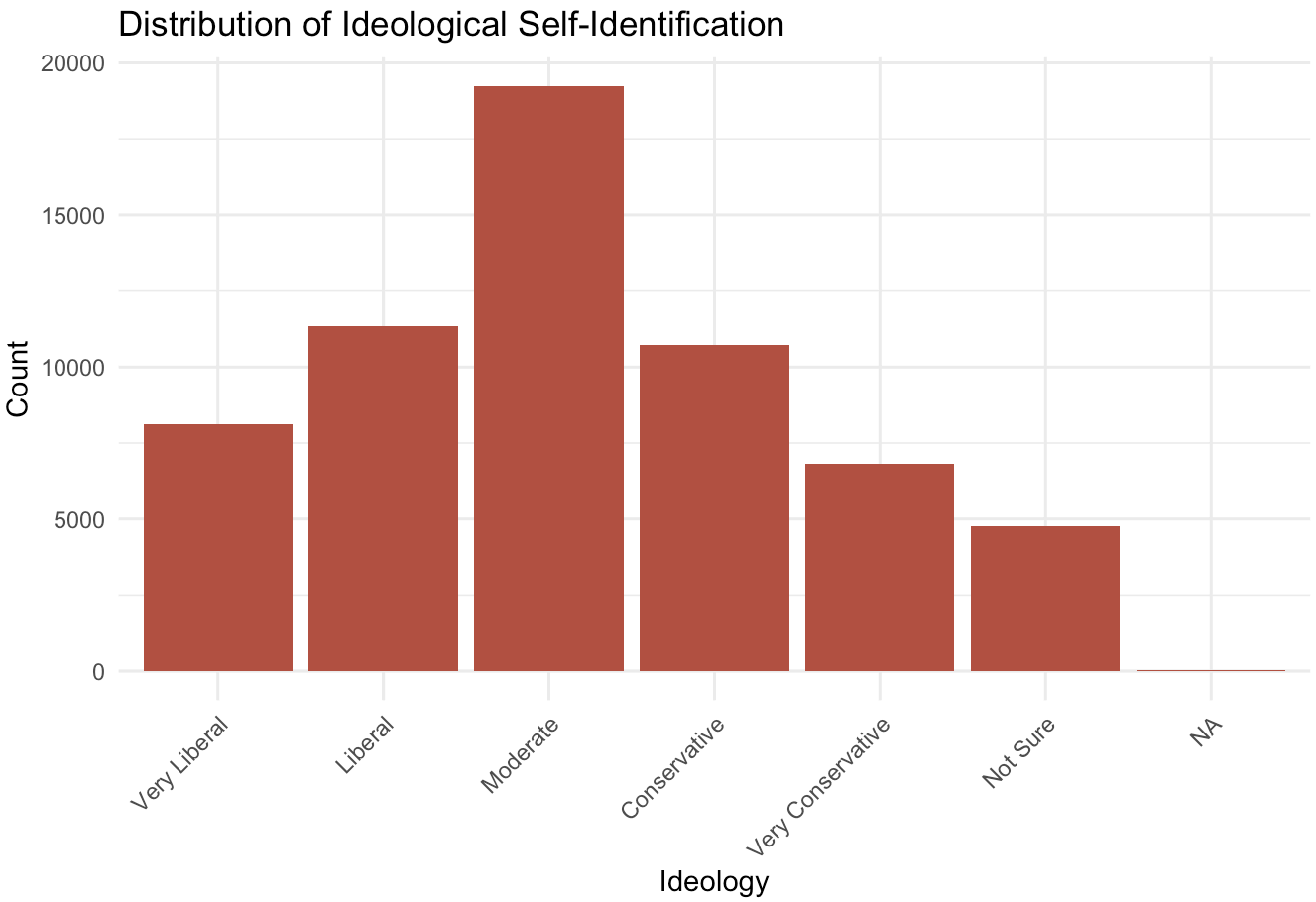
Data Exploration and Analysis

Before diving into our main questions, let's first examine our key variables individually:

```
# Examine ideology distribution
cc_media |>
  filter(year == 2020) |>
  count(ideo5) |>
  mutate(
    ideo5 = haven::as_factor(ideo5),
    pct = n/sum(n) * 100
  ) |>
  arrange(desc(n))
```

```
## # A tibble: 7 × 3
##   ideo5                n     pct
##   <fct>            <int>   <dbl>
## 1 Moderate         19226 31.5
## 2 Liberal          11331 18.6
## 3 Conservative     10716 17.6
## 4 Very Liberal      8132 13.3
## 5 Very Conservative  6803 11.2
## 6 Not Sure           4752  7.79
## 7 <NA>                 40  0.0656
```

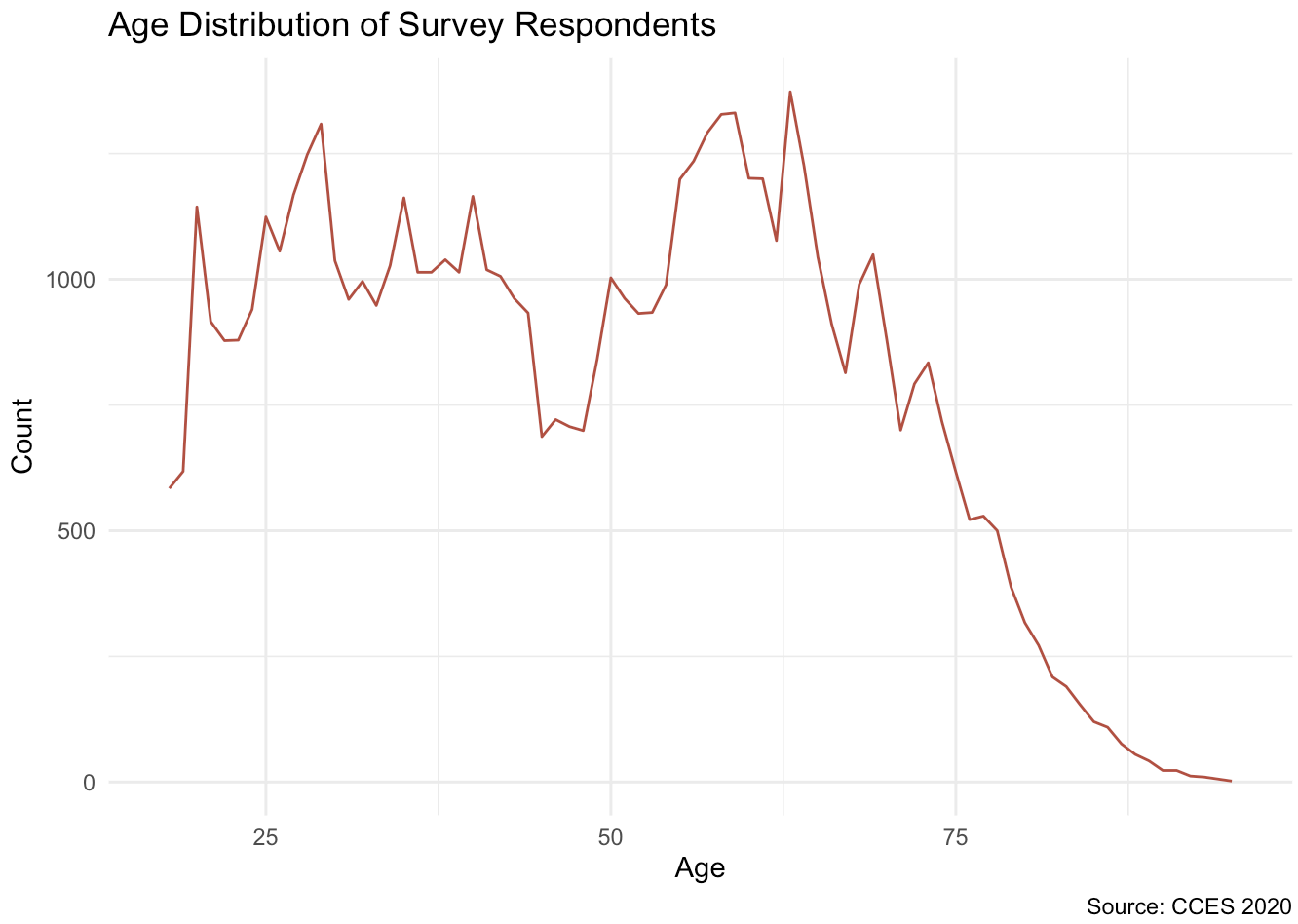
```
# Visualize ideology distribution
cc_media |>
  filter(year == 2020) |>
  mutate(ideo5 = haven::as_factor(ideo5)) |>
  ggplot(aes(x = ideo5)) +
  geom_bar(fill = "#b45346") +
  labs(
    title = "Distribution of Ideological Self-Identification",
    x = "Ideology",
    y = "Count",
    caption = "Source: CCES 2020"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Examine age distribution
cc_media |>
  filter(year == 2020) |>
  summarise(
    mean_age = mean(age, na.rm = TRUE),
    median_age = median(age, na.rm = TRUE),
    sd_age = sd(age, na.rm = TRUE),
    min_age = min(age, na.rm = TRUE),
    max_age = max(age, na.rm = TRUE)
  )
```

```
## # A tibble: 1 × 5
##   mean_age median_age sd_age min_age max_age
##      <dbl>      <dbl>  <dbl>   <dbl>   <dbl>
## 1     48.4         49   17.7      18      95
```

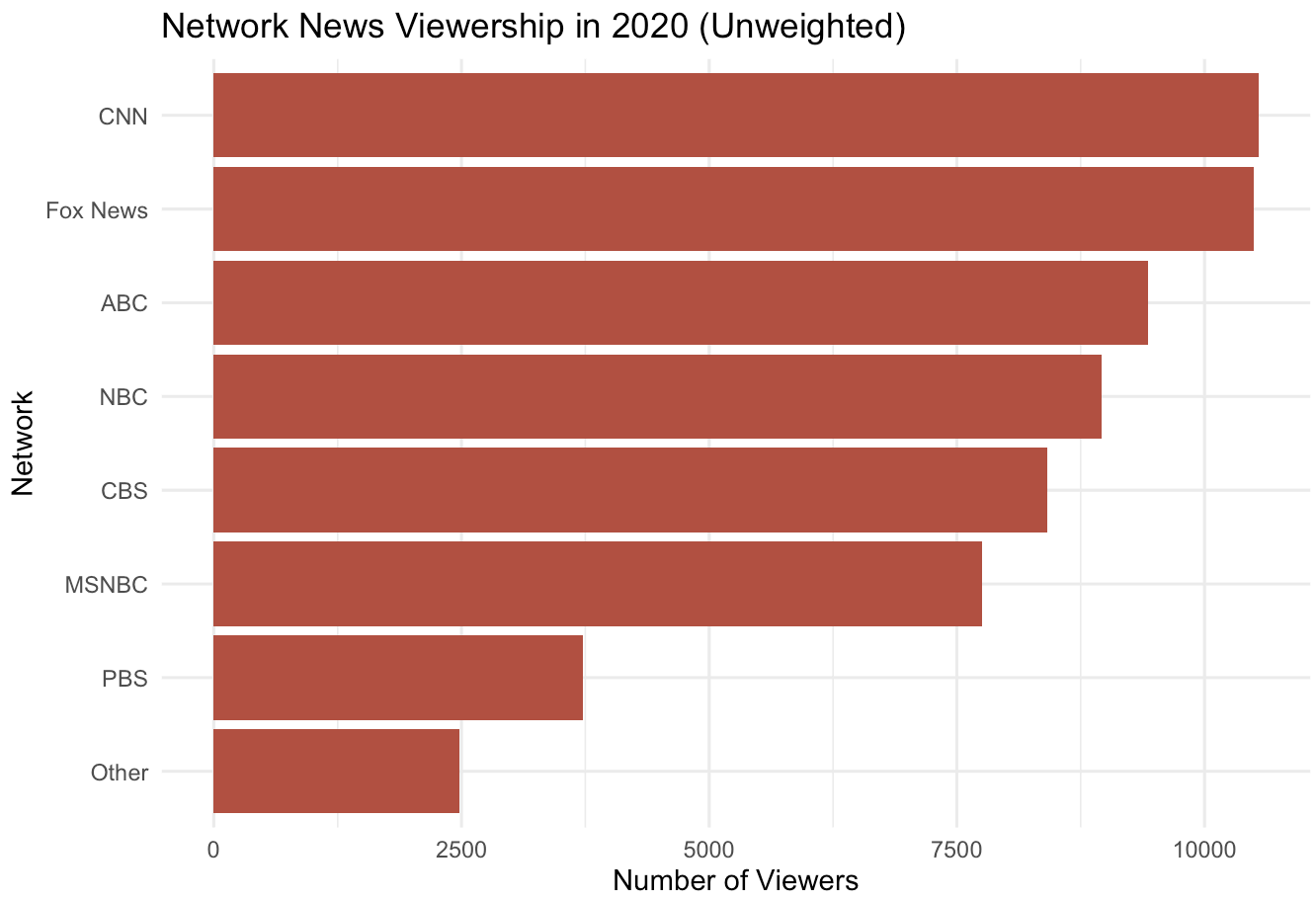
```
# Visualize age distribution
cc_media |>
  filter(year == 2020) |>
  ggplot(aes(x = age)) +
  geom_line(stat = "bin", binwidth = 1, color = "#b45346") +
  labs(
    title = "Age Distribution of Survey Respondents",
    x = "Age",
    y = "Count",
    caption = "Source: CCES 2020"
  ) +
  theme_minimal()
```



```
# Examine overall network viewership
# Note: Uses na.rm=TRUE because some respondents didn't answer all network questions
# This is expected in survey data and shouldn't bias our results as long as
# non-response is relatively random across demographic groups
cc_media |>
  filter(year == 2020) |>
  summarise(across(matches("300b_[1-8]"),
                   ~sum(.x == 1, na.rm = TRUE))) |>
  pivot_longer(everything(),
               names_to = "network",
               values_to = "viewers") |>
  mutate(
    network = clean_network_names(network),
    pct_total = viewers / nrow(filter(cc_media, year == 2020)) * 100
  ) |>
  arrange(desc(viewers))
```

```
## # A tibble: 8 × 3
##   network  viewers pct_total
##   <chr>      <int>     <dbl>
## 1 CNN        10542     17.3
## 2 Fox News   10494     17.2
## 3 ABC         9430     15.5
## 4 NBC         8960     14.7
## 5 CBS         8411     13.8
## 6 MSNBC       7752     12.7
## 7 PBS         3723      6.10
## 8 Other       2483      4.07
```

```r
# Create network viewership plot for 2020
cc_media |>
  filter(year == 2020) |>
  select(matches("300b_[1-8]")) |>
  summarise(across(everything(),
                   ~sum(.x == 1, na.rm = TRUE))) |>
  pivot_longer(everything(),
               names_to = "network",
               values_to = "viewers") |>
  # Clean up network names
  mutate(
    network = clean_network_names(network),
    # Convert to factor and order by viewership
    network = fct_reorder(network, viewers)
  ) |>
  # Create the plot
  ggplot(aes(x = network, y = viewers)) +
  geom_col(fill = "#b45346") +
  coord_flip() + # Horizontal bars
  labs(
    title = "Network News Viewership in 2020 (Unweighted)",
    x = "Network",
    y = "Number of Viewers",
    caption = "Source: CCES 2020"
  ) +
  theme_minimal()
```



Inital Exploration Findings:

1. **Ideology**: Moderates form the largest group (31.5%), with a roughly even split between liberal and conservative leanings.
2. **Age**: Mean age 48.4 years (SD = 17.7), with a typical right-skewed distribution.
3. **Network Viewership**: CNN and Fox News lead (17.3% and 17.2% respectively), followed by broadcast networks (13.8-15.5%).

Research Questions and Analytical Approaches

My analysis examines three key questions about American TV news consumption, each employing different data science skills:

1. **Network Choice and Ideology** (Data Transformation & Visualization)
    - Question: How do ideological preferences relate to news network choice, and to what extent do viewers consume news from multiple sources?
    - Skills: Data reshaping (pivot_longer), grouped summaries, and complex visualizations (stacked bar charts)
    - Variables: Network viewership (categorical), ideology (ordinal)
2. **Age Distribution Analysis** (Statistical Analysis & Density Plots)
    - Question: What is the age distribution of viewers across different networks, and how might this impact the future of TV news?
    - Skills: Density estimation, statistical summaries, and overlapping distribution plots
    - Variables: Age (continuous), network viewership (categorical)

3. **Policy Preference Alignment** (Statistical Modeling & Heat Maps)
    - Question: How closely do viewers' policy preferences align with their chosen news sources?
    - Skills: Linear regression modeling, correlation analysis, and heat map visualization
    - Variables: Policy positions (categorical), network viewership (categorical), ideology (ordinal)

Q1: Network Viewership and Ideology

Crosstabs ideology and media consumption 2020:

```
# Create ideology and media consumption analysis
cc_media_ideo <- cc_media |>
  filter(year == 2020) |>
  # Select network preferences and ideology
  select(matches("300b_[1-8]"), ideo5) |>
  # Reshape network data to long format
  pivot_longer(
    cols = matches("300b_"),
    names_to = "network",
    values_to = "watches"
  ) |>
  # Clean up network names
  mutate(
    network = clean_network_names(network),
    # Convert ideo5 to factor
    ideo5 = haven::as_factor(ideo5)
  ) |>
  # Filter for those who watch (watches == 1)
  filter(watches == 1) |>
  # Create cross-tabulation
  count(network, ideo5) |>
  # Convert to percentages within each network
  group_by(network) |>
  mutate(pct = n / sum(n) * 100) |>
  ungroup()

cc_media_ideo
```

```
## # A tibble: 55 × 4
##    network ideo5                n     pct
##    <chr>   <fct>            <int>   <dbl>
##  1 ABC     Very Liberal      1210 12.8
##  2 ABC     Liberal           2238 23.7
##  3 ABC     Moderate          3582 38.0
##  4 ABC     Conservative      1329 14.1
##  5 ABC     Very Conservative  608  6.45
##  6 ABC     Not Sure           454  4.81
##  7 ABC     <NA>                 9  0.0954
##  8 CBS     Very Liberal      1046 12.4
##  9 CBS     Liberal           1931 23.0
## 10 CBS     Moderate          3252 38.7
## # ℹ 45 more rows
```

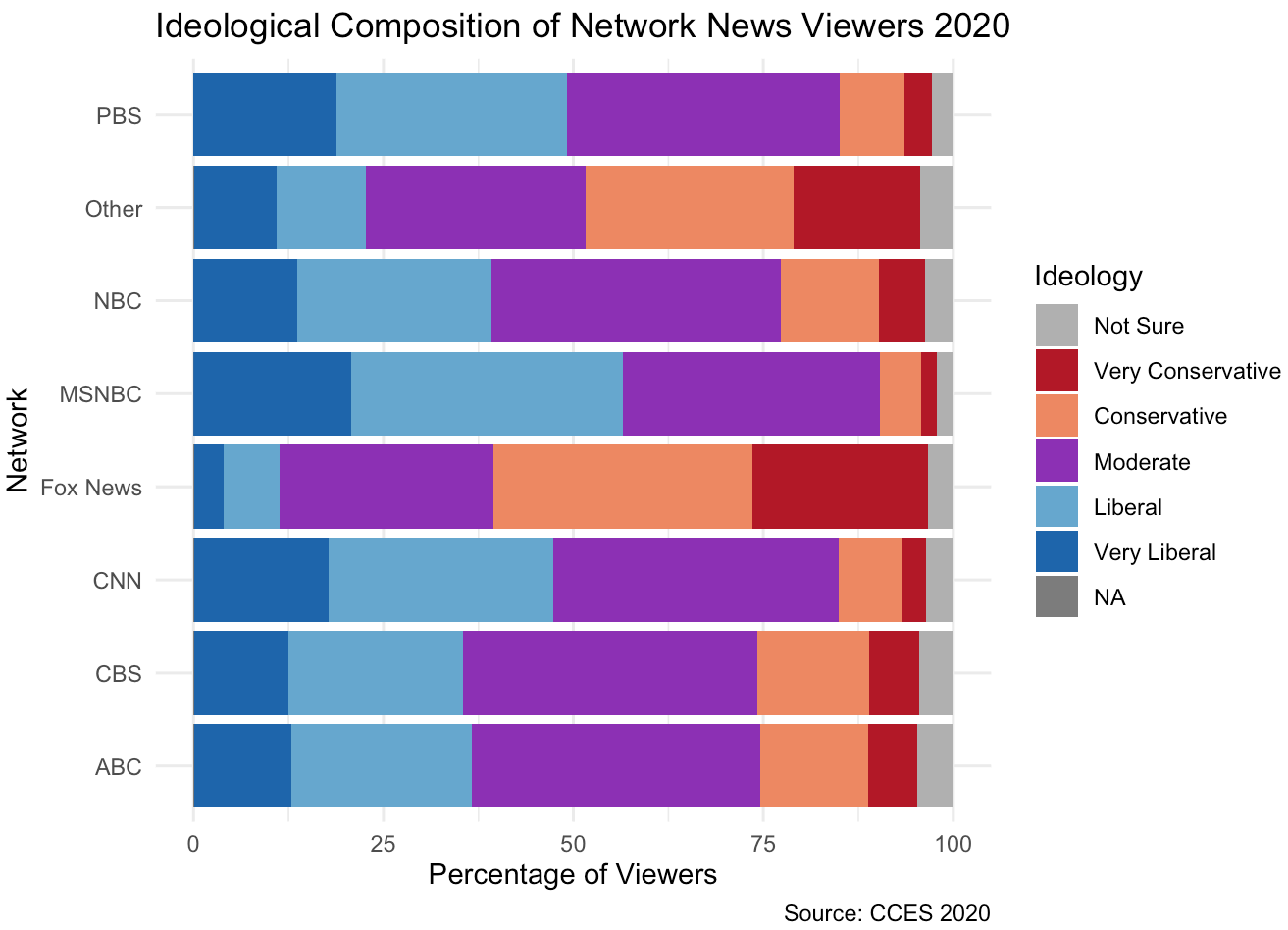
Formatted Crosstabs

```
# Create formatted table
cc_media_ideo |>
  # Round percentages
  mutate(pct = round(pct, 1)) |>
  # Pivot wider for easier reading
  pivot_wider(
    names_from = ideo5,
    values_from = c(n, pct)
  ) |>
  # Arrange by total viewership
  arrange(desc(rowSums(across(starts_with("n_")), na.rm = TRUE)))
```

```
## # A tibble: 8 × 15
##   network  `n_Very Liberal` n_Liberal n_Moderate n_Conservative `n_Very Conservative`
`n_Not Sure`  n_NA
##   <chr>               <int>     <int>      <int>          <int>                 <int>
<int> <int>
## 1 CNN                 1872      3111       3963            874                   335
384     3
## 2 Fox News             413       773       2955           3576                  2419
352     6
## 3 ABC                 1210      2238       3582           1329                   608
454     9
## 4 NBC                 1221      2291       3411           1157                   548
331     1
## 5 CBS                 1046      1931       3252           1244                   556
374     8
## 6 MSNBC               1606      2770       2629            415                   165
165     2
## 7 PBS                  701      1128       1337            320                   133
104    NA
## 8 Other                270       292        718            678                   415
108     2
## # ℹ 7 more variables: `pct_Very Liberal` <dbl>, pct_Liberal <dbl>, pct_Moderate <dbl
>, pct_Conservative <dbl>,
## #   `pct_Very Conservative` <dbl>, `pct_Not Sure` <dbl>, pct_NA <dbl>
```

Ideological composition of network news viewers bar chart 2020

```
# Create stacked bar chart
ggplot(cc_media_ideo, aes(x = network, y = pct, fill = fct_rev(ideo5))) +
  geom_col() +
  coord_flip() +
  scale_fill_manual(values = c(
    "Not Sure" = "grey70",
    "NA" = "grey90",
    "Very Liberal" = "#2166ac",      # Dark blue
    "Liberal" = "#67a9cf",           # Light blue
    "Moderate" = "#8f34b6",          # Purple
    "Conservative" = "#ef8a62",      # Light red
    "Very Conservative" = "#b2182b"  # Dark red
  )) +
  labs(
    title = "Ideological Composition of Network News Viewers 2020",
    x = "Network",
    y = "Percentage of Viewers",
    fill = "Ideology",
    caption = "Source: CCES 2020"
  ) +
  theme_minimal()
```



1. **Strong ideologues show different viewing patterns**
    - Very Conservatives heavily favor Fox News (23.1% of Fox viewers)
    - Very Liberals favor MSNBC and CNN (20.7% and 17.8% of their viewers)
    - Both groups have lower viewership of centrist networks
2. **Moderates dominate broadcast network viewership**
    - Moderates make up ~38% of ABC, CBS, and NBC viewers
    - Moderates also represent significant portions of CNN (37.6%) and PBS (35.9%)
    - Even Fox News has substantial moderate viewership (28.2%)
3. **Clear ideological network preferences**
    - Fox News viewers skew conservative (57.2% Conservative/Very Conservative)
    - MSNBC viewers skew liberal (56.4% Liberal/Very Liberal)
    - Broadcast networks (ABC, CBS, NBC) maintain more balanced viewership

This suggests network preferences align with ideological leanings, though most networks maintain a significant moderate audience base. Broadcast networks appear most successful at reaching across ideological lines.

How many people watch multiple networks in 2020?

```
# Check multiple network viewing
cc_media |>
  filter(year == 2020) |>
  select(matches("300b_[1-8]")) |>
  # Count how many networks each person watches (where value == 1)
  mutate(
    networks_watched = rowSums(across(everything(), ~.x == 1), na.rm = TRUE)
  ) |>
  # Summarize the distribution
  count(networks_watched) |>
  # Calculate percentages
  mutate(pct = n/sum(n) * 100)
```

```
## # A tibble: 9 × 3
##   networks_watched     n     pct
##              <dbl> <int>   <dbl>
## 1                0 35208 57.7
## 2                1  8917 14.6
## 3                2  7266 11.9
## 4                3  4425  7.25
## 5                4  2558  4.19
## 6                5  1395  2.29
## 7                6   796  1.30
## 8                7   392  0.643
## 9                8    43  0.0705
```

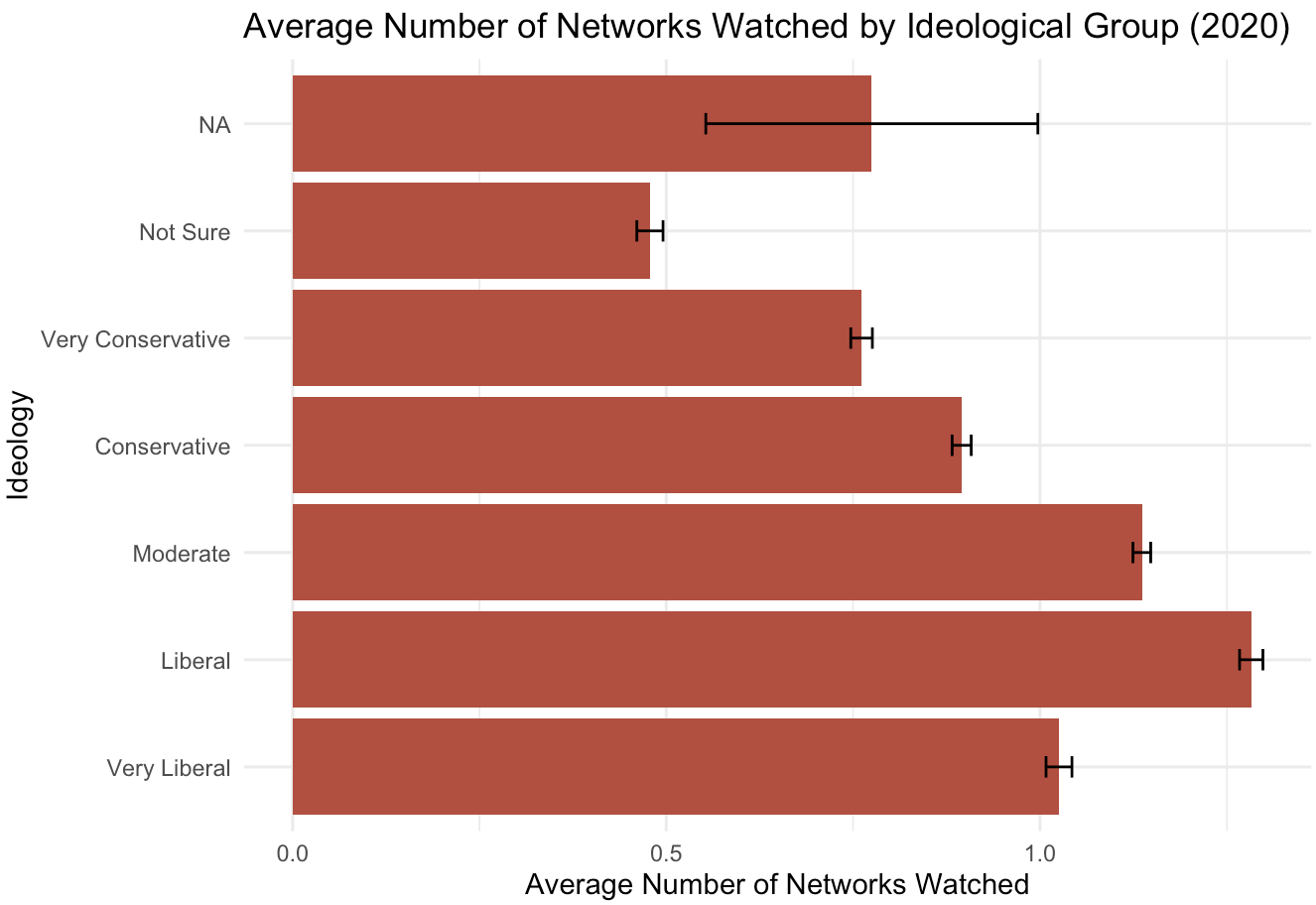
Cross tab of relationship between ideology and number of networks watched in 2020

```
# Analyze ideology vs number of networks watched
cc_media |>
  filter(year == 2020) |>
  select(matches("300b_[1-8]"), ideo5) |>
  # Count networks watched and convert ideology to factor
  mutate(
    networks_watched = rowSums(across(matches("300b_"), ~.x == 1), na.rm = TRUE),
    ideo5 = haven::as_factor(ideo5)
  ) |>
  # Group by ideology and calculate mean networks watched
  group_by(ideo5) |>
  summarise(
    n = n(),  # number of respondents
    mean_networks = mean(networks_watched, na.rm = TRUE),
    median_networks = median(networks_watched, na.rm = TRUE),
    zero_networks = mean(networks_watched == 0, na.rm = TRUE) * 100,  # % watching zero
    multiple_networks = mean(networks_watched > 1, na.rm = TRUE) * 100  # % watching mul
tiple
  ) |>
  # Remove NA ideology if present
  filter(!is.na(ideo5))
```

```
## # A tibble: 6 × 6
##   ideo5                n mean_networks median_networks zero_networks multiple_networ
ks
##   <fct>            <int>         <dbl>           <dbl>         <dbl>             <db
l>
## 1 Very Liberal      8132         1.03               0          59.0              2
8.1
## 2 Liberal          11331         1.28               0          50.9              3
6.2
## 3 Moderate         19226         1.14               0          56.4              3
0.9
## 4 Conservative     10716         0.895              0          55.9              2
4.5
## 5 Very Conservative  6803        0.761              0          57.7              1
9.2
## 6 Not Sure          4752         0.478              0          81.0              1
2.6
```

bar chart of above FOR 2020

```
# Create bar chart of average networks watched by ideology
cc_media |>
  filter(year == 2020) |>
  select(matches("300b_[1-8]"), ideo5) |>
  mutate(
    networks_watched = rowSums(across(matches("300b_"), ~.x == 1), na.rm = TRUE),
    ideo5 = haven::as_factor(ideo5)
  ) |>
  # Calculate mean networks watched per ideology
  group_by(ideo5) |>
  summarise(
    avg_networks = mean(networks_watched, na.rm = TRUE),
    se = sd(networks_watched, na.rm = TRUE) / sqrt(n())  # standard error for error bars
  ) |>
  # Create the plot
  ggplot(aes(x = ideo5, y = avg_networks)) +
  geom_col(fill = "#b45346") +
  geom_errorbar(aes(ymin = avg_networks - se, ymax = avg_networks + se),
                width = 0.2) +
  labs(
    title = "Average Number of Networks Watched by Ideological Group (2020)",
    x = "Ideology",
    y = "Average Number of Networks Watched",
    caption = "Source: CCES 2020"
  ) +
  theme_minimal() +
  coord_flip()
```



1. **Strong ideologues show different viewing patterns**
    - Very Conservatives average 0.76 networks
    - Very Liberals average 1.03 networks
    - Both groups have lower multiple-network viewing rates (19.2% and 28.1% respectively)
2. **Liberals and moderates consume more diverse sources**
    - Liberals watch most networks (mean = 1.28)
    - Moderates follow closely (mean = 1.14)
    - Liberals most likely to watch multiple networks (36.2%)
3. **Overall low TV news consumption**
    - ~51-59% watch zero networks across most ideological groups
    - Exception: “Not Sure” group shows even lower engagement (81.0% watch none)

This suggests ideological polarization may relate to more limited news source selection, with liberals showing the highest engagement with multiple sources. Traditional TV news networks have low viewership across all groups.

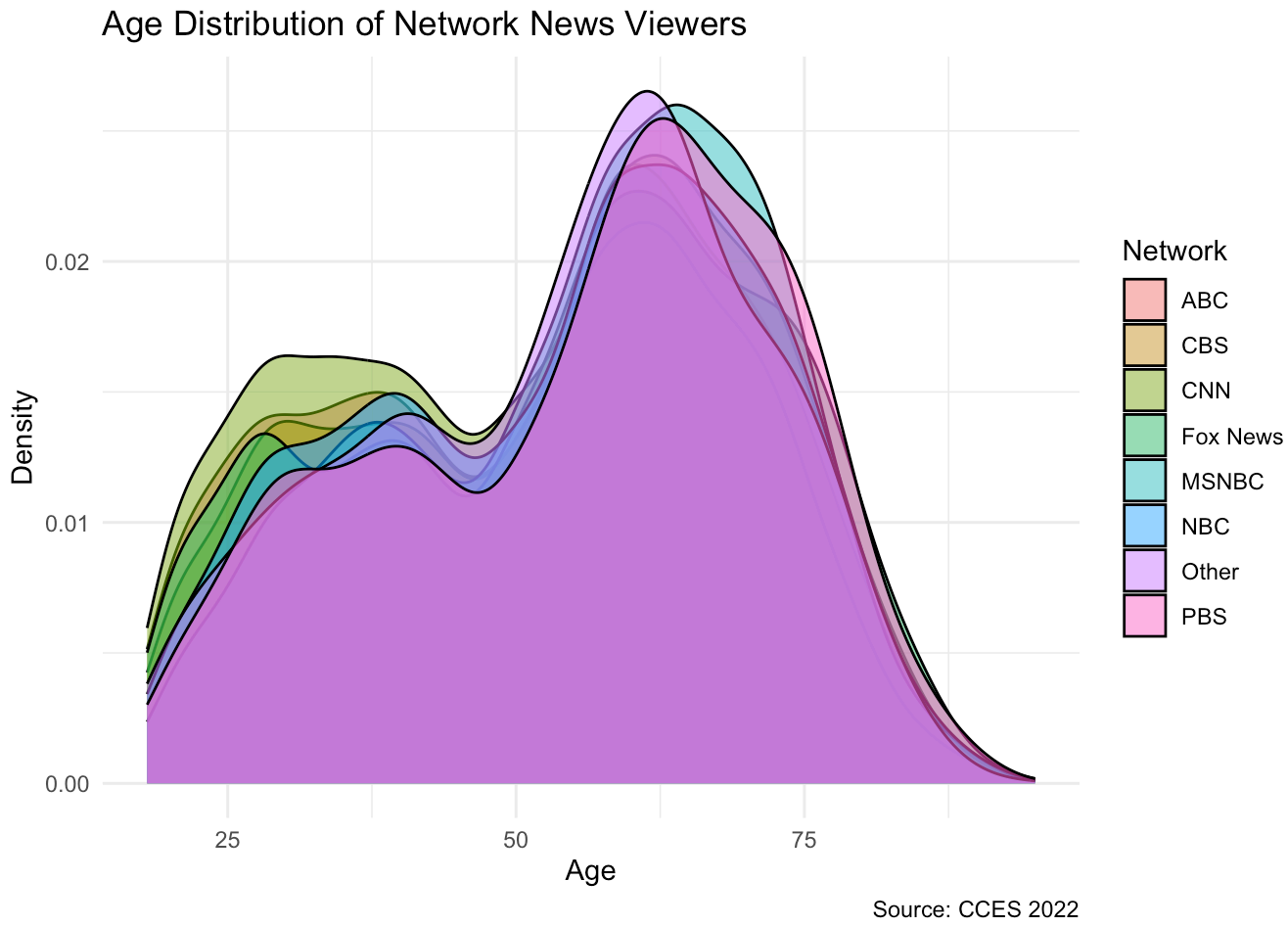
Q2: Age Distribution

```
# Analyze age by network viewership
cc_media |>
  filter(year == 2020) |>
  select(matches("300b_[1-8]"), age) |>
  # Reshape to long format
  pivot_longer(
    cols = matches("300b_"),
    names_to = "network",
    values_to = "watches"
  ) |>
  # Clean network names and group broadcast networks
  mutate(
    network = clean_network_names(network),
  ) |>
  # Filter for viewers
  filter(watches == 1) |>
  # Calculate age statistics by network
  group_by(network) |>
  summarise(
    mean_age = mean(age, na.rm = TRUE),
    median_age = median(age, na.rm = TRUE),
    n = n()
  ) |>
  arrange(desc(mean_age))
```

```
## # A tibble: 8 × 4
##   network  mean_age median_age     n
##   <chr>       <dbl>      <dbl> <int>
## 1 PBS          55.8         59  3723
## 2 MSNBC        55.5         58  7752
## 3 Other        54.2         57  2483
## 4 NBC          54.1         57  8960
## 5 Fox News     53.9         57 10494
## 6 CBS          53.6         57  8411
## 7 ABC          52.4         56  9430
## 8 CNN          50.3         52 10542
```

age groups and visualize the distribution

```
# Create age distribution plot
cc_media |>
  filter(year == 2020) |>
  select(matches("300b_[1-8]"), age) |>
  pivot_longer(
    cols = matches("300b_"),
    names_to = "network",
    values_to = "watches"
  ) |>
  mutate(
    network = clean_network_names(network),
  ) |>
  filter(watches == 1) |>
  ggplot(aes(x = age, fill = network)) +
  geom_density(alpha = 0.5) +
  labs(
    title = "Age Distribution of Network News Viewers",
    x = "Age",
    y = "Density",
    fill = "Network",
    caption = "Source: CCES 2022"
  ) +
  theme_minimal()
```



Analysis of Age and Network News Viewership:

1. CNN has notably younger audience (mean age 50, median 52)
2. PBS and MSNBC have oldest viewers (mean age 56)
3. Broadcast Networks (NBC, CBS, ABC) range from 52-54 mean age
4. Fox News viewers align with overall trend (mean 54, median 57)
5. All networks skew older, with median ages ranging from 52-59

This suggests traditional TV news, regardless of network, primarily reaches older Americans, with CNN being somewhat more successful at attracting younger viewers.

Q3: Policy Preferences

What is the correlation between respondents' policy preferences and their media source preferences? Additionally, how does this correlation compare to the relationship between their policy preferences and ideology?

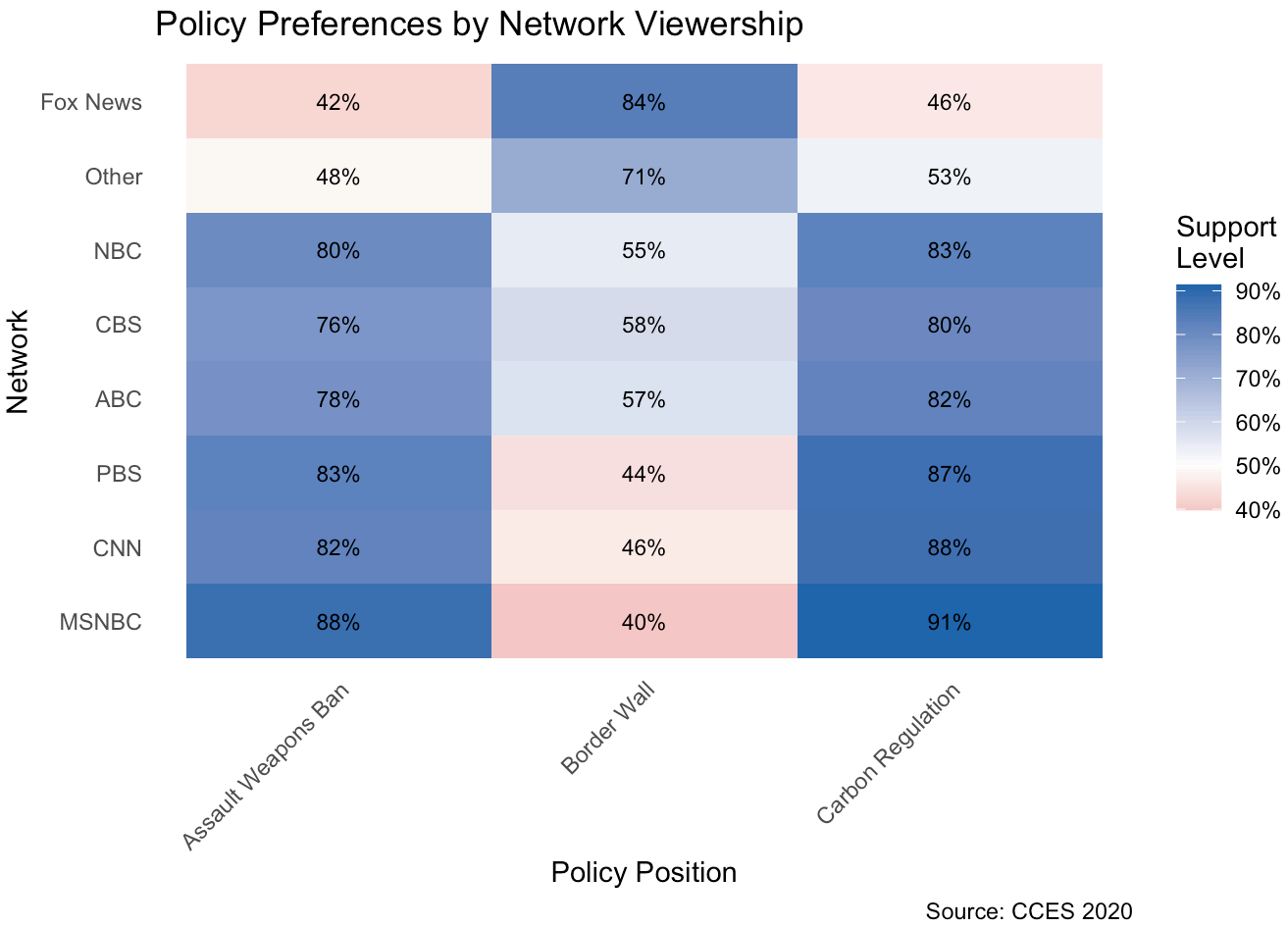
```
# Create a summary of policy preferences by network viewership
network_policy_summary <- cc_media_policy |>
  select(matches("300b_[1-8]"), enviro_carbon, immig_border, guns_assaultban) |>
  # Reshape to long format for network viewing
  pivot_longer(
    cols = matches("300b_"),
    names_to = "network",
    values_to = "watches"
  ) |>
  filter(watches == 1) |>  # Keep only viewers
  mutate(
    network = clean_network_names(network)
  ) |>
  group_by(network) |>
  summarise(
    carbon_support = mean(enviro_carbon == 1, na.rm = TRUE),
    border_support = mean(immig_border == 1, na.rm = TRUE),
    assault_ban_support = mean(guns_assaultban == 1, na.rm = TRUE),
    n = n()
  ) |>
  arrange(desc(n))

network_policy_summary
```

```
## # A tibble: 8 × 5
##   network  carbon_support border_support assault_ban_support     n
##   <chr>             <dbl>          <dbl>               <dbl> <int>
## 1 CNN               0.877          0.463               0.822 10542
## 2 Fox News          0.460          0.841               0.425 10494
## 3 ABC               0.818          0.572               0.777  9430
## 4 NBC               0.829          0.547               0.799  8960
## 5 CBS               0.803          0.583               0.765  8411
## 6 MSNBC             0.914          0.398               0.879  7752
## 7 PBS               0.874          0.444               0.828  3723
## 8 Other             0.528          0.712               0.485  2483
```

```r
# Reshape data for heatmap
network_policy_long <- network_policy_summary |>
  select(-n) |>
  pivot_longer(
    cols = ends_with("_support"),
    names_to = "policy",
    values_to = "support"
  ) |>
  mutate(
    # Clean up policy names for display
    policy = case_when(
      policy == "carbon_support" ~ "Carbon Regulation",
      policy == "border_support" ~ "Border Wall",
      policy == "assault_ban_support" ~ "Assault Weapons Ban"
    ),
    network = fct_relevel(network,
      "MSNBC", "CNN", "PBS", "ABC", "CBS", "NBC", "Other", "Fox News")
  )

# Create heatmap
ggplot(network_policy_long,
       aes(x = policy, y = network, fill = support)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "#b2182b",    # Republican red
    mid = "white",
    high = "#2166ac",   # Democratic blue
    midpoint = 0.5,
    labels = scales::percent
  ) +
  # Add text labels
  geom_text(aes(label = scales::percent(support, accuracy = 1)),
            color = "black", size = 3) +
  labs(
    title = "Policy Preferences by Network Viewership",
    x = "Policy Position",
    y = "Network",
    fill = "Support\nLevel",
    caption = "Source: CCES 2020"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  )
```



Source: CCES 2020

1. **Clear Network-Policy Alignment**
    - MSNBC and Fox News viewers show starkly opposing policy preferences
    - MSNBC viewers strongly support climate and gun control (91%, 88%)
    - Fox News viewers strongly support border wall (84%)
2. **Broadcast Network Consistency**
    - ABC, CBS, and NBC viewers show similar preferences
    - Generally align with CNN/MSNBC but less strongly
    - Support levels: ~80% for carbon regulation and assault ban, ~55% for border wall
3. **Policy Polarization**
    - Border wall shows greatest viewer polarization (84% Fox vs 40% MSNBC)
    - Carbon regulation and assault weapon ban show similar patterns
    - PBS viewers align closely with CNN on all issues

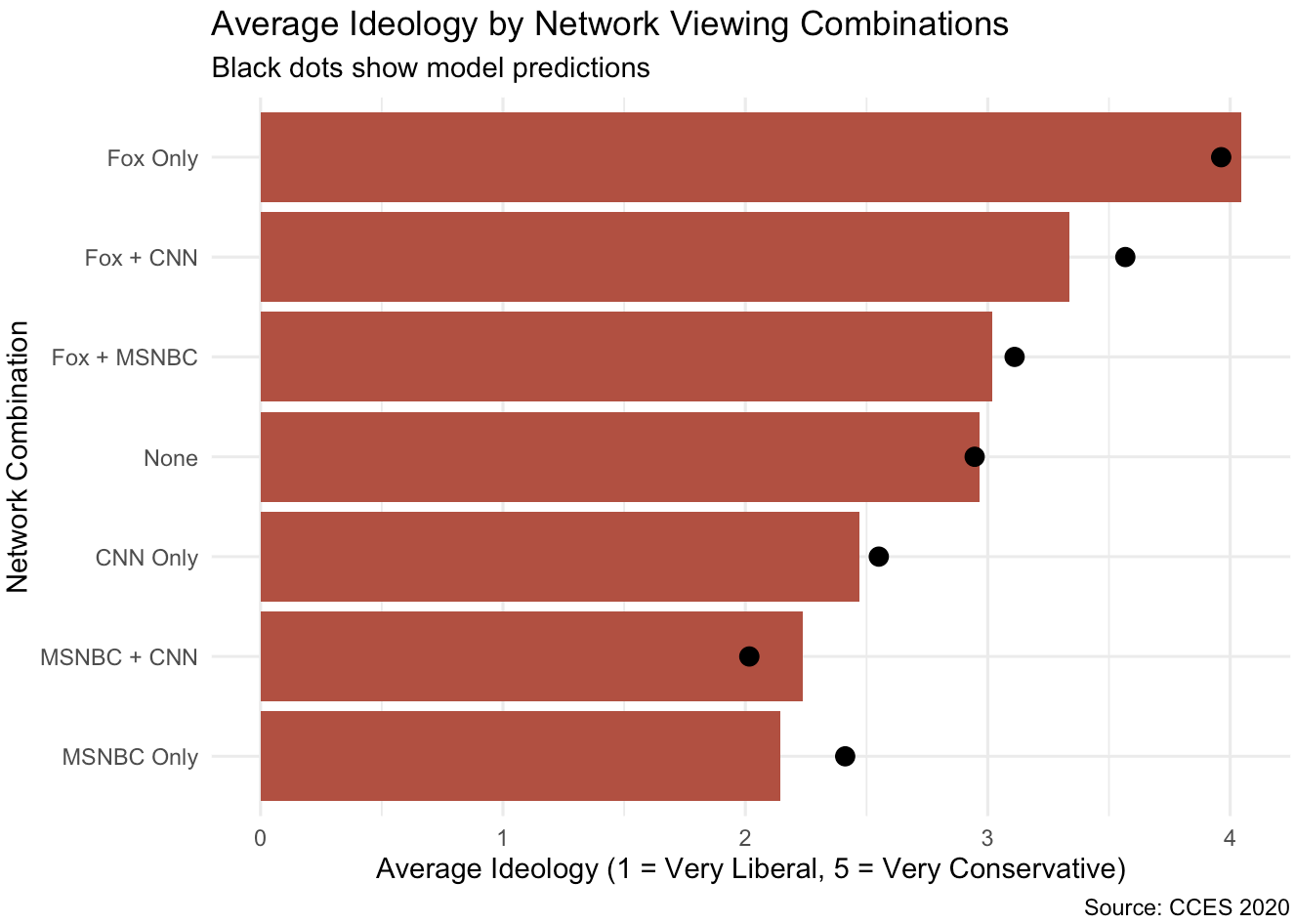
```r
# Create dummy variables for watching each network
cc_media_policy <- cc_media_policy |>
  mutate(
    watches_fox = `300b_5` == 1,
    watches_msnbc = `300b_6` == 1,
    watches_cnn = `300b_4` == 1
  )

# Regression of ideology on network choice
mod_ideology <- lm(ideo5 ~ watches_fox + watches_msnbc + watches_cnn,
                   data = cc_media_policy)
summary(mod_ideology)
```

```
##
## Call:
## lm(formula = ideo5 ~ watches_fox + watches_msnbc + watches_cnn,
##     data = cc_media_policy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9630 -0.9460  0.0370  0.5881  3.9836
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.94600    0.01191  247.37   <2e-16 ***
## watches_foxTRUE    1.01705    0.01441   70.57   <2e-16 ***
## watches_msnbcTRUE -0.53412    0.01609  -33.19   <2e-16 ***
## watches_cnnTRUE   -0.39551    0.01480  -26.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.099 on 25793 degrees of freedom
##   (35203 observations deleted due to missingness)
## Multiple R-squared:  0.2795, Adjusted R-squared:  0.2794
## F-statistic:  3336 on 3 and 25793 DF,  p-value: < 2.2e-16
```

1. **Strong Network-Ideology Relationship**
    - Fox News viewing associated with +1.02 points on ideology scale (more conservative)
    - MSNBC viewing associated with -0.53 points (more liberal)
    - CNN viewing associated with -0.40 points (more liberal)
    - All effects highly significant ($p < 0.001$)
2. **Model Fit**
    - $R^2 = 0.28$, suggesting network choice explains 28% of ideological variance
    - Baseline (intercept) ideology is 2.95 (slightly conservative)

```r
# Create visualization with proper handling of missing values
cc_media_policy |>
  # First filter to only the complete cases used in the regression
  filter(!is.na(ideo5), !is.na(watches_fox), !is.na(watches_msnbc), !is.na(watches_cnn))
|>
  mutate(
    predicted_ideology = predict(mod_ideology),
    network_combo = case_when(
      watches_fox & !watches_msnbc & !watches_cnn ~ "Fox Only",
      !watches_fox & watches_msnbc & !watches_cnn ~ "MSNBC Only",
      !watches_fox & !watches_msnbc & watches_cnn ~ "CNN Only",
      watches_fox & watches_msnbc ~ "Fox + MSNBC",
      watches_fox & watches_cnn ~ "Fox + CNN",
      watches_msnbc & watches_cnn ~ "MSNBC + CNN",
      watches_fox & watches_msnbc & watches_cnn ~ "All Three",
      TRUE ~ "None"
    )
  ) |>
  group_by(network_combo) |>
  summarise(
    avg_ideology = mean(ideo5, na.rm = TRUE),
    pred_ideology = mean(predicted_ideology, na.rm = TRUE),
    n = n()
  ) |>
  filter(n > 100) |>  # Filter for combinations with sufficient data
  ggplot(aes(x = reorder(network_combo, avg_ideology), y = avg_ideology)) +
  geom_col(fill = "#b45346") +
  geom_point(aes(y = pred_ideology), color = "black", size = 3) +
  coord_flip() +
  labs(
    title = "Average Ideology by Network Viewing Combinations",
    subtitle = "Black dots show model predictions",
    x = "Network Combination",
    y = "Average Ideology (1 = Very Liberal, 5 = Very Conservative)",
    caption = "Source: CCES 2020"
  ) +
  theme_minimal()
```



Conclusion

This analysis of CCES data reveals several important patterns in American TV news consumption. The findings demonstrate significant relationships between viewers' ideological preferences, age, and network choices.

1. While ideological sorting exists in news viewership (e.g., Fox News viewers skewing conservative, MSNBC liberal), most networks maintain substantial moderate audiences. Even Fox News and MSNBC draw roughly 28-30% of their viewers from self-identified moderates.
2. TV news viewership skews older across all networks. CNN attracts the youngest audience with a median age of 52, while other networks range from 54-59. This age skew raises questions about the future viability of TV news as a medium, especially as younger generations increasingly turn to digital platforms for news.
3. Strong correlations exist between viewers' policy preferences and their chosen news sources, particularly on issues like immigration and climate change. However, broadcast networks (ABC, CBS, NBC) show more ideologically diverse viewership than cable news networks.

Limitations of this analysis include: - Reliance on self-reported viewing habits - Online-only survey methodology may underrepresent certain populations - Focus on traditional TV news excludes emerging digital media sources - Cross-sectional nature of the analysis limits causal inference

Future research could: - Examine how streaming and social media affect news consumption patterns - Track individual-level changes in viewing habits over time - Investigate the relationship between local news viewership and political attitudes - Compare U.S. patterns with those in other countries