

## investigation-US-election-2016

```
library(conflicted)
  conflicts_prefer(dplyr::filter, dplyr::lag)
library(tidyverse)
library(R6)
library(formatR)
library(lubridate)
```

```
allstates = polls$state |>
  unique() |>
  sort()
polls = polls |>
  mutate(
    interval = startdate %--% enddate,
    middate = ymd(startdate + (enddate - startdate)/2),
    .keep = "unused",
    .after = "state"
  ) |>
  mutate(
    grade = grade |>
      factor(
        levels = c(
          "A+",
          "A",
          "A-",
          "B+",
          "B",
          "B-",
          "C+",
          "C",
          "C-",
          "D+",
          "D",
          "D-",
          NA
        )
      ),
    .keep = "unused",
    .before = "samplesize"
  )
```

```
finalresults = tibble(
  state = allstates,
  clinton = c(34.36, 36.55, 45.13, 33.65, 61.73, 48.16, 54.57, 53.09, 90.86, 47.82, 45.64, 62.22, 27.4),
  trump = c(62.08, 51.28, 48.67, 60.57, 31.62, 43.25, 40.93, 41.71, 4.09, 49.02, 50.77, 30.04, 59.26),
  johnson = c(2.09, 5.88, 4.13, 2.65, 3.37, 5.18, 2.96, 3.33, 1.58, 2.20, 3.05, 3.72, 4.10, 3.79, 4.9)
```

```
mcmullin = c(NA, NA, 0.68, 1.17, 0.28, 1.04, 0.13, 0.16, NA, NA, 0.32, NA, 6.73, 0.21, NA, 0.79, 0.1)
)
```

```
# helps filter for candidate name in results
```

```
candidatenames = c("clinton", "trump", "johnson", "mcmullin")
```

```
candidate = R6Class(
```

```
  classname = "candidate",
```

```
  public = list(
```

```
    name = "character",
```

```
    opponents = "character",
```

```
    colour = "character",
```

```
    polls = "tbl_df",
```

```
    finalresults = "tbl_df",
```

```
  initialize = function(name, colour){
```

```
    self$name = name
```

```
    self$colour = colour
```

```
    self$opponents = candidatenames[candidatenames != self$name] # the list of other candidates
```

```
    self$polls = polls |>
```

```
      select(!contains(self$opponents)) |> # filter polls to only this candidate
```

```
      rename(
```

```
        rawpolls = starts_with("rawpoll"),
```

```
        adjpolls = starts_with("adjpoll")
```

```
      ) # rename columns for use in candidateplot functions
```

```
    self$finalresults = finalresults |>
```

```
      select(!contains(self$opponents))
```

```
  } # end of initialize
```

```
) # end of list
```

```
)
```

```
clinton = candidate$new(name = "clinton", colour = "blue")
```

```
trump = candidate$new(name = "trump", colour = "red")
```

```
johnson = candidate$new(name = "johnson", colour = "orange")
```

```
mcmullin = candidate$new(name = "mcmullin", colour = "purple")
```

```
candidates = list(clinton, trump, johnson, mcmullin)
```

```
rm(candidatenames) # only needed for candidate construction
```

```
densepollsbegin = function(thestate, thedate) { # for graphing purposes
```

```
  # calculate the earliest date after thedate (inclusive) where there are three
```

```
  # consecutive polls in short succession (within two months)
```

```
  statepolls = polls |>
```

```
    filter(state == thestate, middate >= thedate) |>
```

```
    arrange(middate)
```

```
  earliestdate = statepolls$middate |>
```

```
    min()
```

```
  earlydate = earliestdate
```

```
  while (nrow(statepolls) > 2) {
```

```
    statepolls = statepolls[-1,]
```

```
    threemos = interval(
```

```
      start = earlydate,
```

```
      end = earlydate %m+% months(2, abbreviate = FALSE)
```

```

    )
    if (
      threemos |>
      int_overlaps(
        statepolls$interval[which.min(sapply(statepolls$interval, int_start))]
      )
    ) {
      temppolls = statepolls[-1,]
      if (
        threemos |>
        int_overlaps(
          temppolls$interval[which.min(sapply(temppolls$interval, int_start))]
        )
      ) {
        return(earlydate)
        break
      }
    }
    earlydate = statepolls$middat |>
      min()
  }
  return(earliestdate) # failed to find three such polls; just graph all polls
}

earliestdate = ymd("2016-05-01")
finaldate = ymd("2016-11-08")

```

```

candidateplotraw = function(thestate, thecandidate, thefirstdate) {
  statepolls = thecandidate$polls |>
    filter(state == thestate)
  if ( # in this state, not all polls for this candidate are NA
    any(!is.na(statepolls$rawpolls))
  ) {
    datevsraw = aes(
      x = statepolls$middat,
      y = statepolls$rawpolls
    )
    return(
      last_plot() + geom_point(
        mapping = datevsraw,
        colour = thecandidate$colour,
        na.rm = TRUE
      ) + geom_smooth(
        mapping = datevsraw,
        colour = thecandidate$colour,
        fill = thecandidate$colour,
        alpha = 0.4,
        na.rm = TRUE
      ) + geom_segment(
        mapping = aes(
          x = thefirstdate,
          y = filter(thecandidate$finalresults, state == thestate)[1,2] |>
            as.numeric(),

```

```

        xend = finaldate,
        yend = filter(thecandidate$finalresults, state == thestate)[1,2] |>
            as.numeric()
    ),
    colour = thecandidate$colour,
    linewidth = 1.0
)
) # end of return
} # end of if
else return(last_plot()) # catch the case where the candidate was not polled in the given state
}

candidateplotadj = function(thestate, thecandidate, thefirstdate) {
    statepolls = thecandidate$polls |>
        filter(state == thestate)
    if ( # in this state, not all polls for this candidate are NA
        any(!is.na(statepolls$adjpolls))
    ) {
        datevsadj = aes(
            x = statepolls$midddate,
            y = statepolls$adjpolls
        )
        return(
            last_plot() + geom_point(
                mapping = datevsadj,
                colour = thecandidate$colour,
                na.rm = TRUE
            ) + geom_smooth(
                mapping = datevsadj,
                colour = thecandidate$colour,
                fill = thecandidate$colour,
                alpha = 0.4,
                na.rm = TRUE
            ) + geom_segment(
                mapping = aes(
                    x = thefirstdate,
                    y = filter(thecandidate$finalresults, state == thestate)[1,2] |>
                        as.numeric(),
                    xend = finaldate,
                    yend = filter(thecandidate$finalresults, state == thestate)[1,2] |>
                        as.numeric()
                ),
                colour = thecandidate$colour,
                linewidth = 1.0
            )
        )
    ) # end of return
} # end of if
else return(last_plot()) # catch the case where the candidate was not polled in the given state
}

```

```

stateplotraw = function(thestate) {
    plot = ggplot()
    firstdatetoplot = thestate |>

```

```

    densepollsbegin(earliestdate)
  for (cand in candidates) plot = candidateplotraw(theState, cand, firstdateToplot)
  stateRaw = theState |>
    paste("Raw Polls", sep = " - ")
  firstdateToplot = theState |>
    densepollsbegin(earliestdate)

  return(
    plot + geom_vline(xintercept = finaldate) + labs(
      title = stateRaw,
      x = "Date",
      y = "Voting Intent (%)"
    ) + xlim(firstdateToplot, finaldate)
  )
}

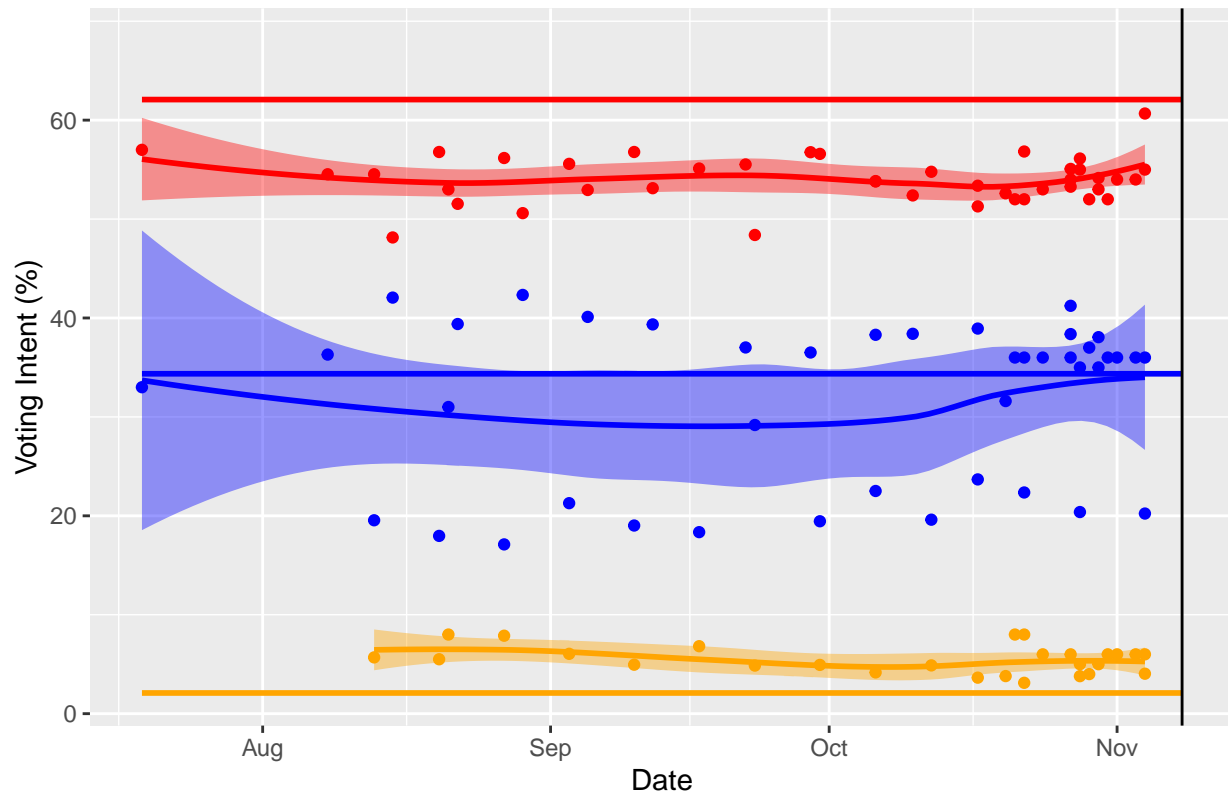
stateplotadj = function(theState) {
  plot = ggplot()
  firstdateToplot = theState |>
    densepollsbegin(earliestdate)
  for (cand in candidates) plot = candidateplotadj(theState, cand, firstdateToplot)
  stateadj = theState |>
    paste("Adjusted Polls", sep = " - ")

  return(
    plot + geom_vline(xintercept = finaldate) + labs(
      title = stateadj,
      x = "Date",
      y = "Adjusted Voting Intent (%)"
    ) + xlim(firstdateToplot, finaldate)
  )
}

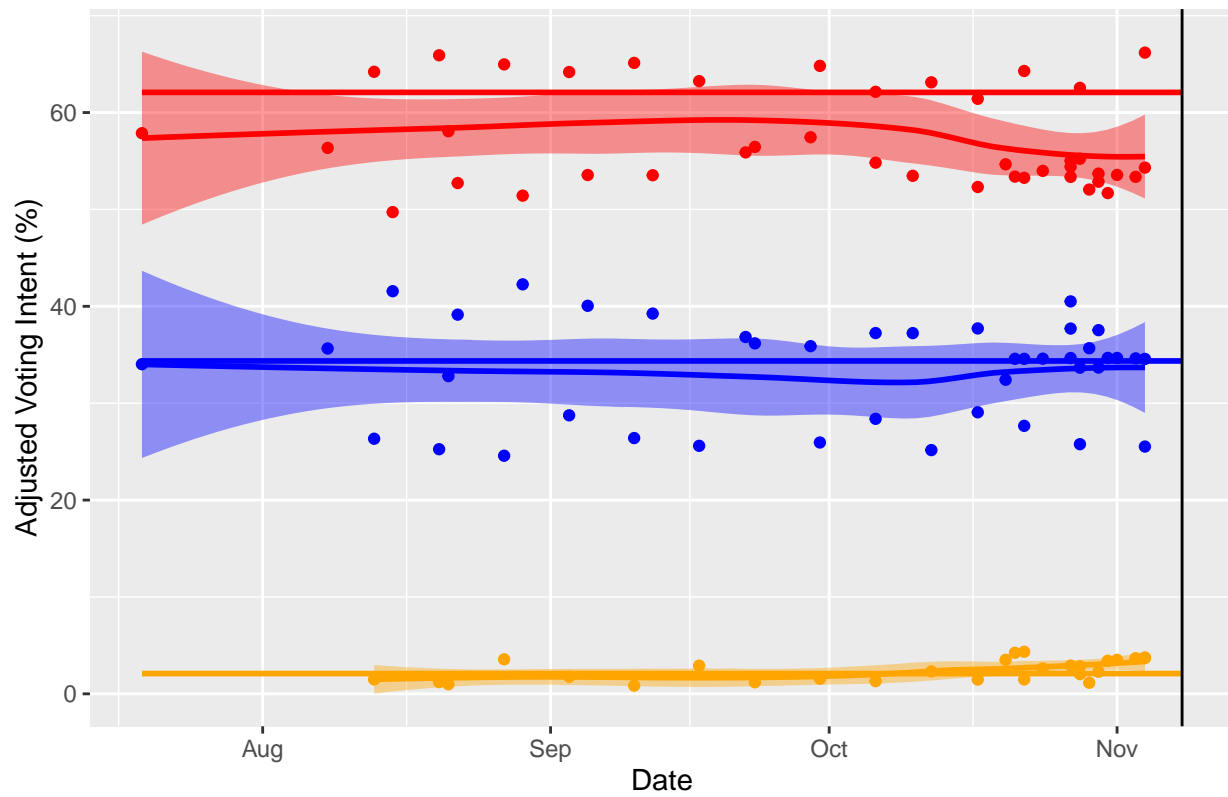
for (theState in allStates) {
  stateplotraw(theState) |>
    print()
  stateplotadj(theState) |>
    print()
}

```

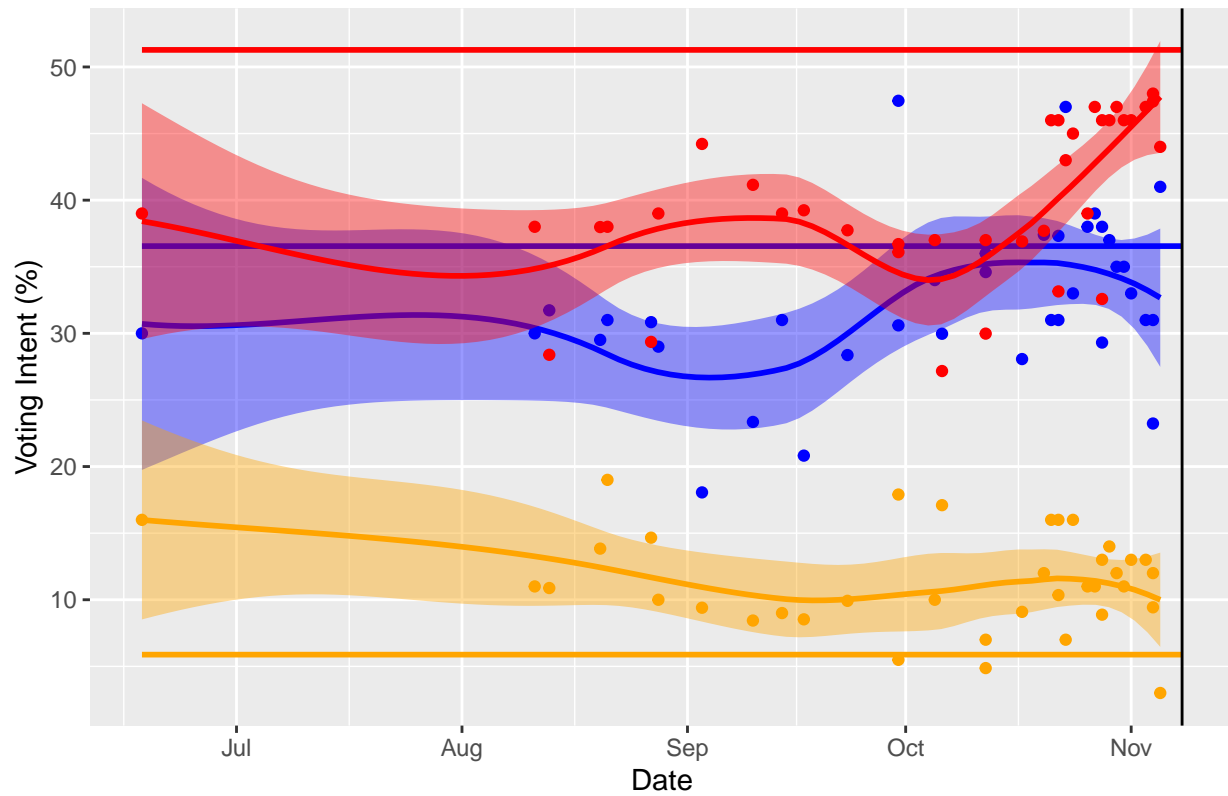
Alabama – Raw Polls



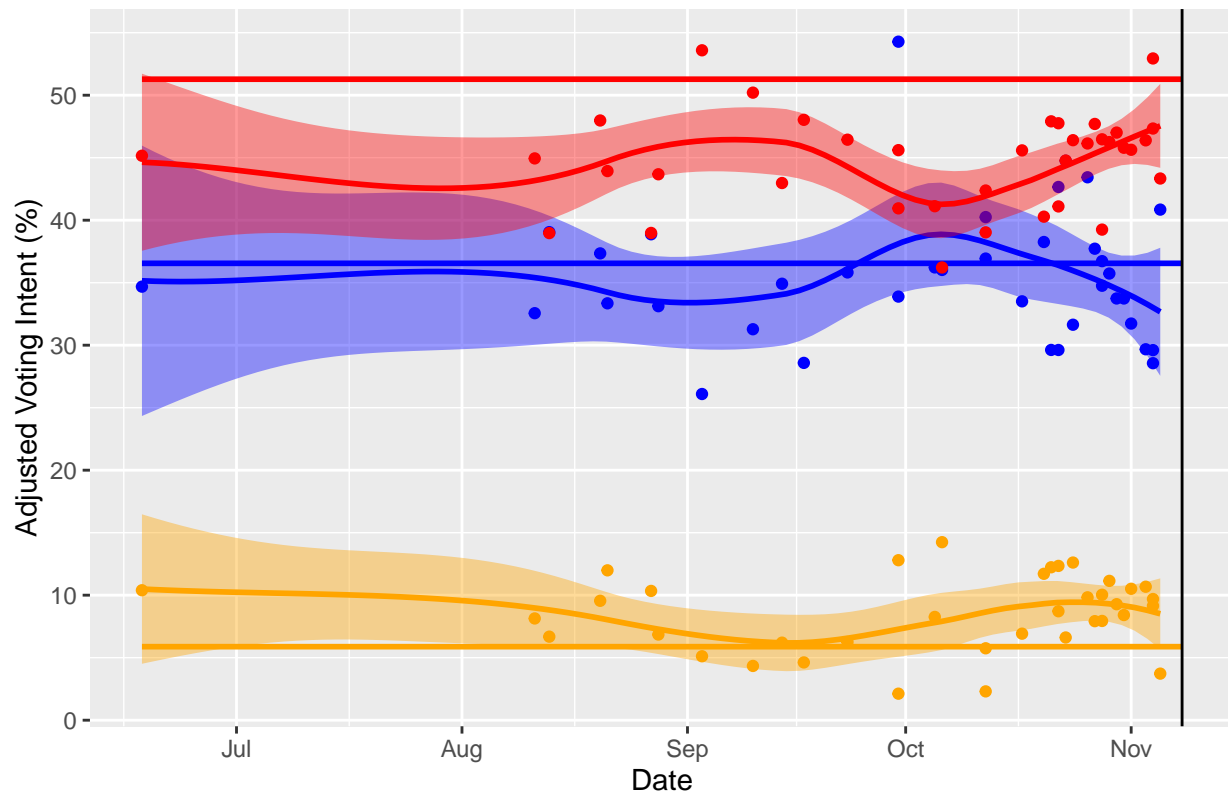
Alabama – Adjusted Polls



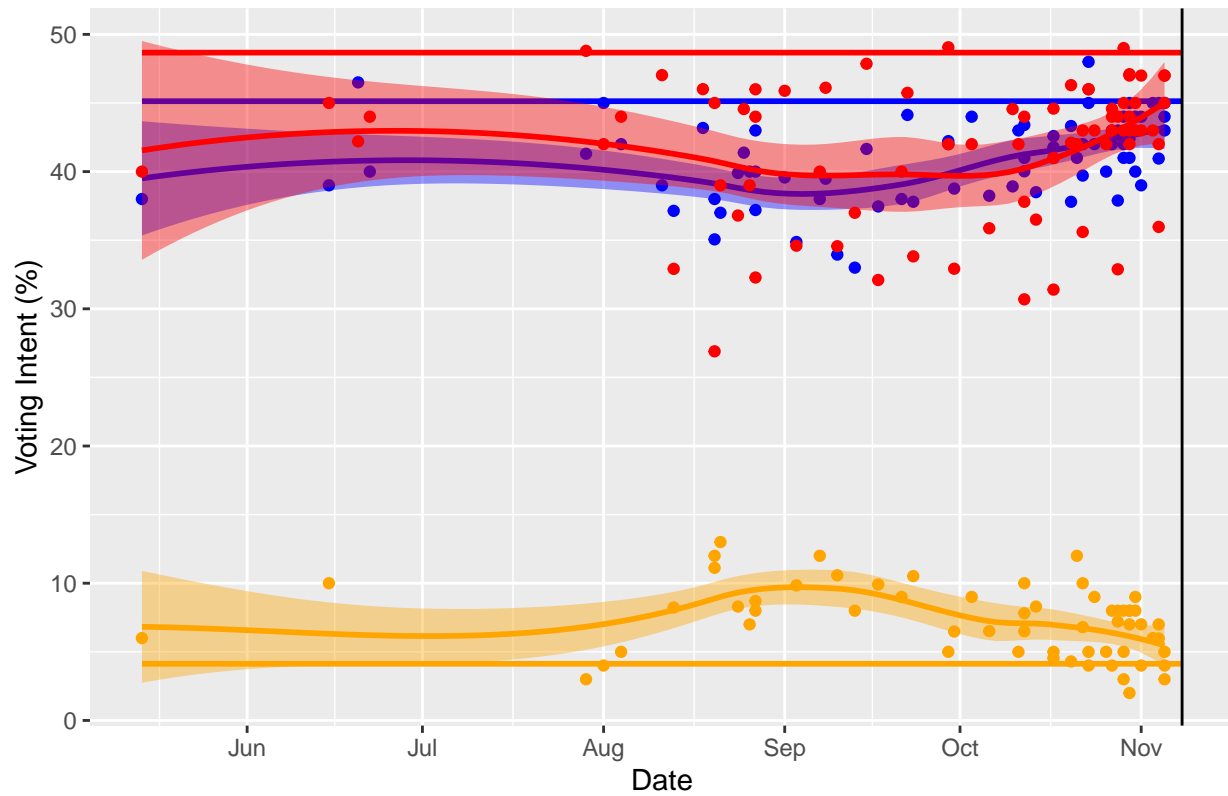
Alaska – Raw Polls



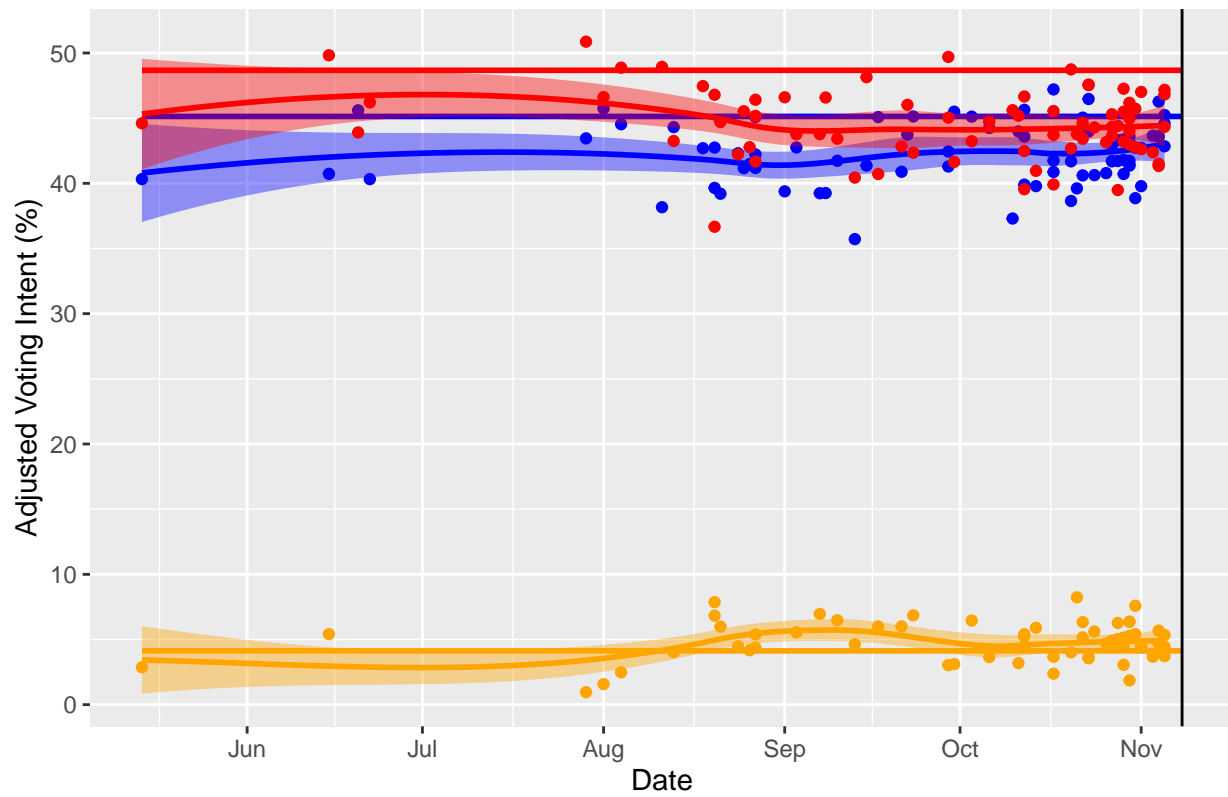
Alaska – Adjusted Polls



Arizona – Raw Polls

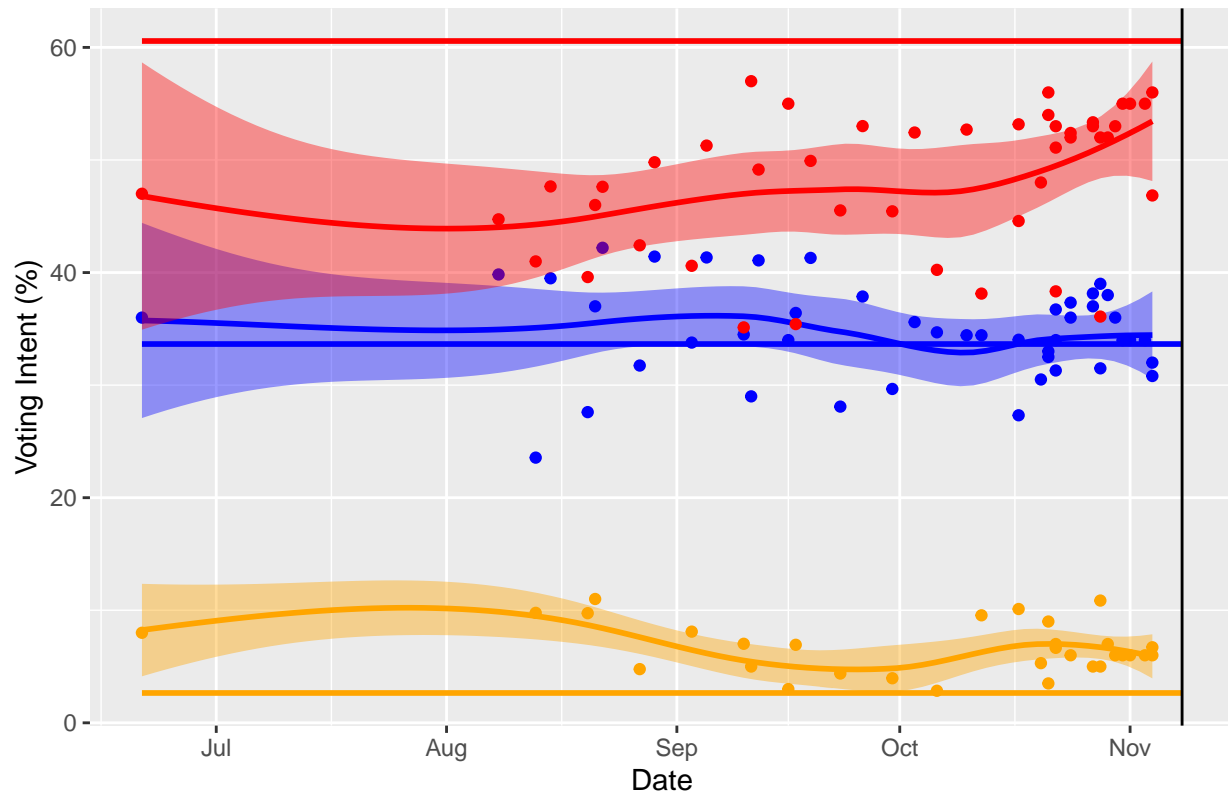


Arizona – Adjusted Polls

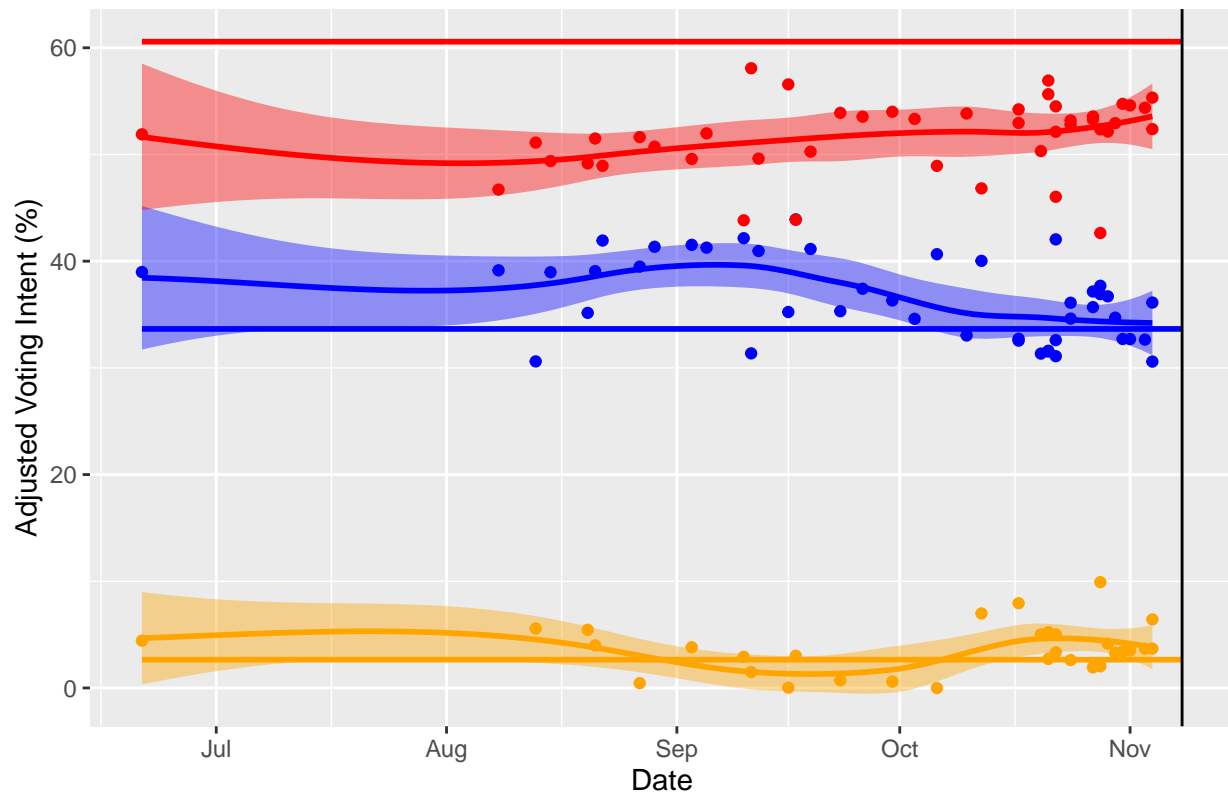




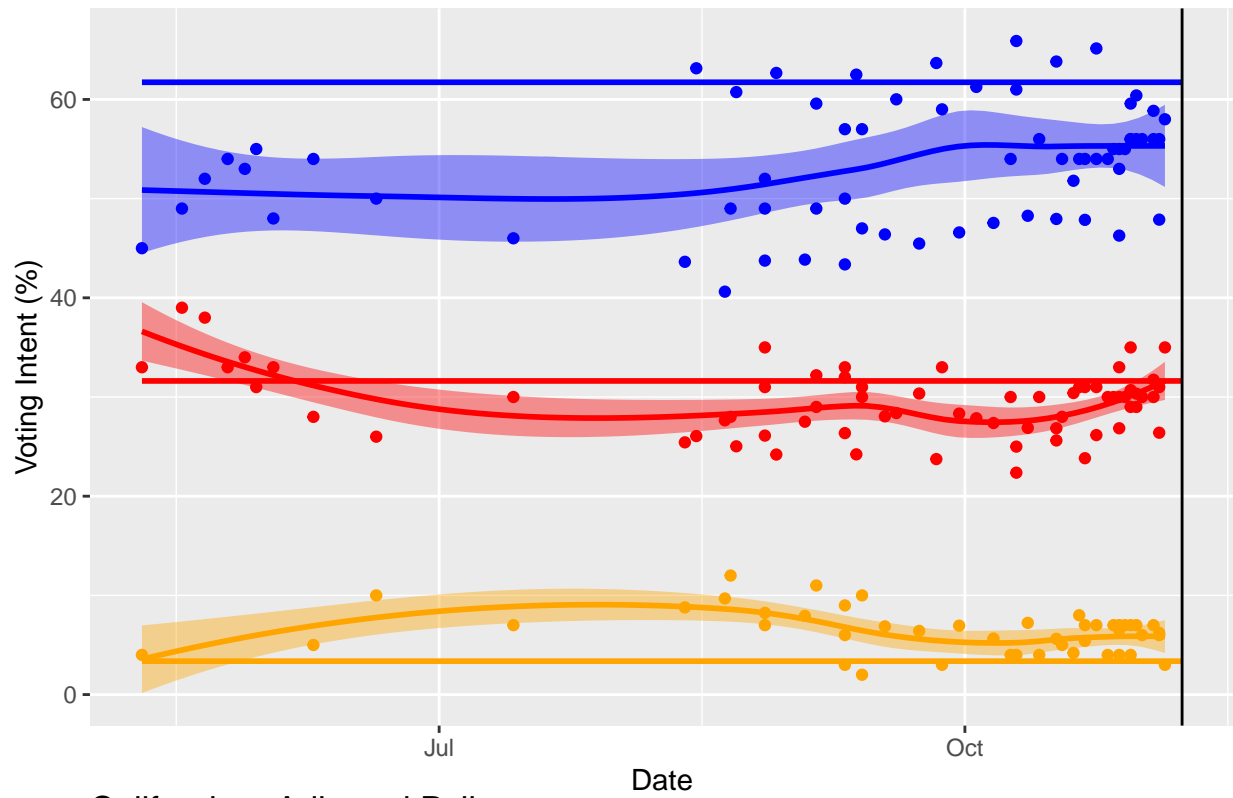
Arkansas – Raw Polls



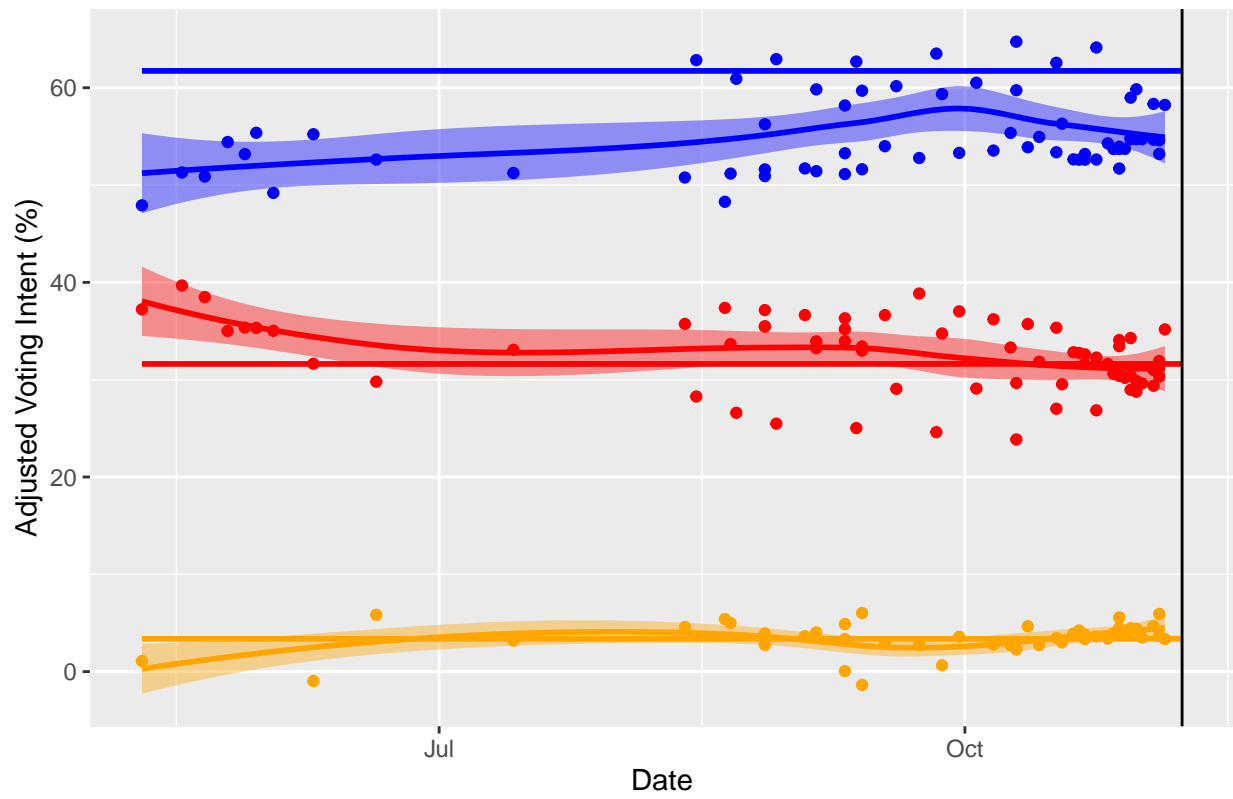
Arkansas – Adjusted Polls



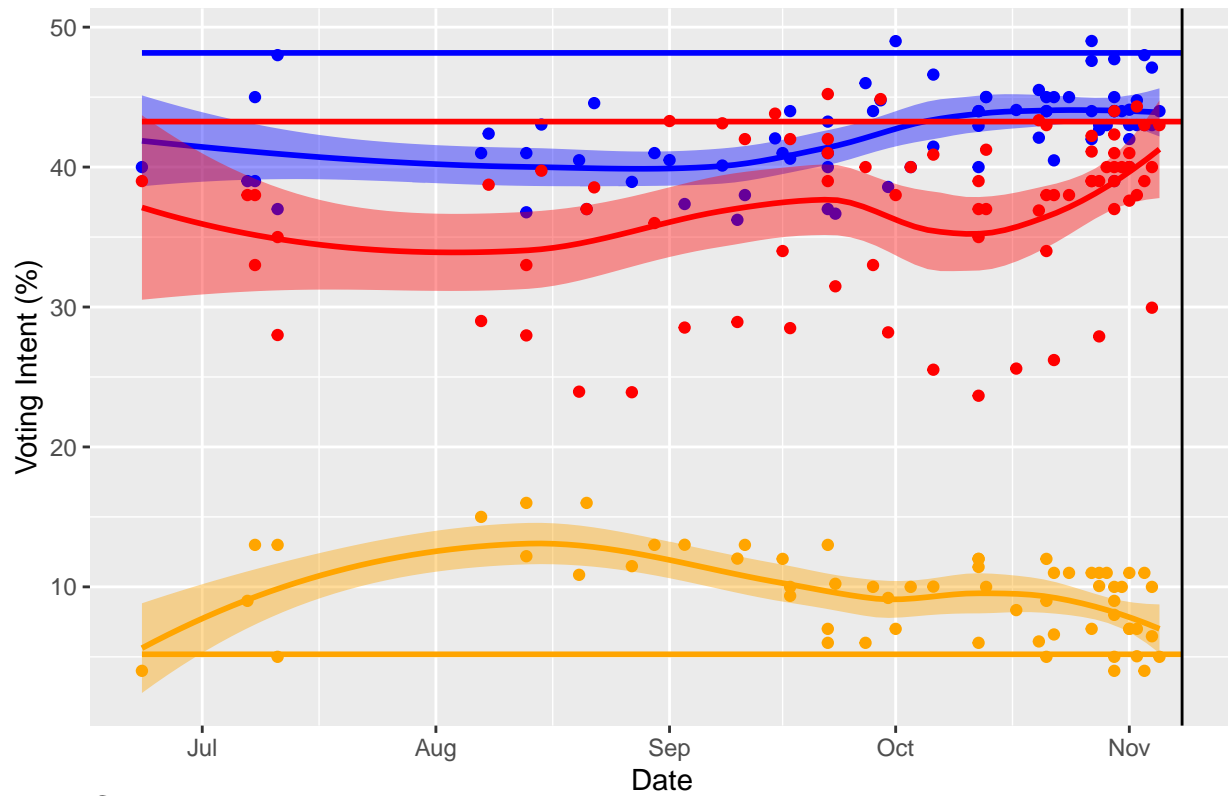
California – Raw Polls



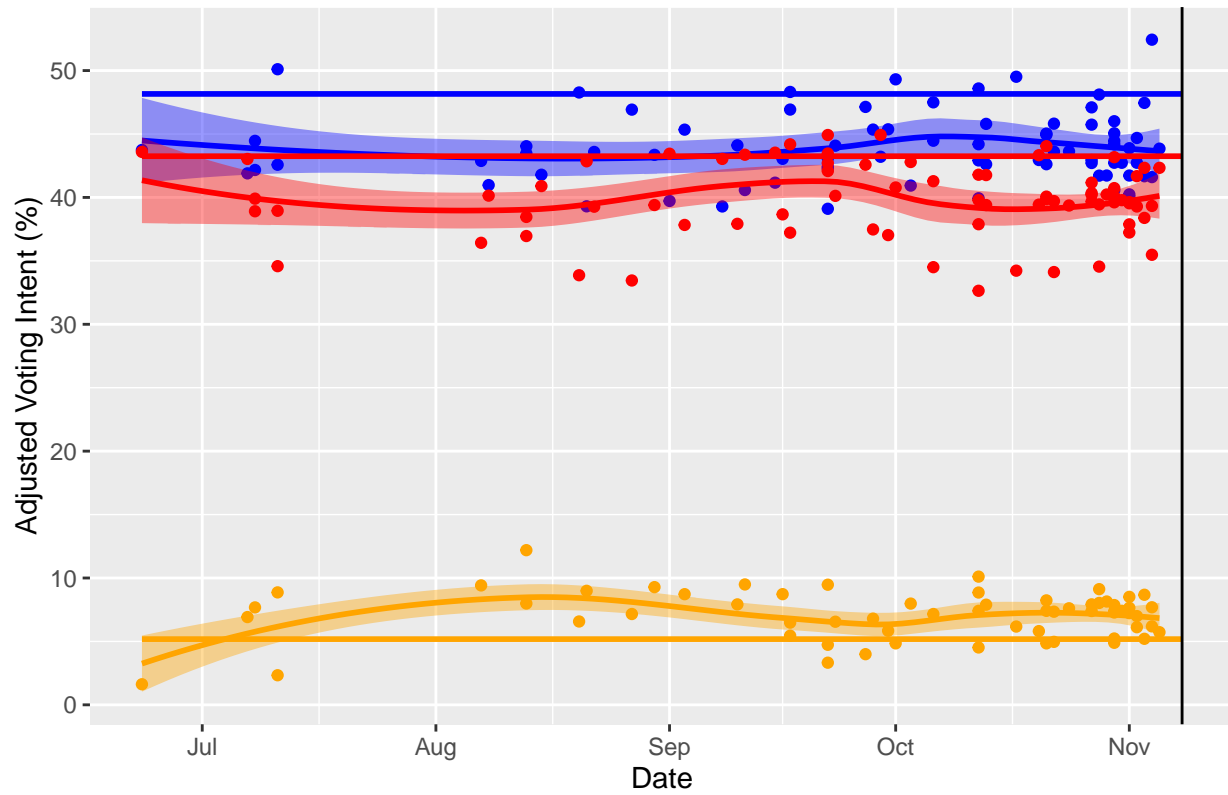
California – Adjusted Polls



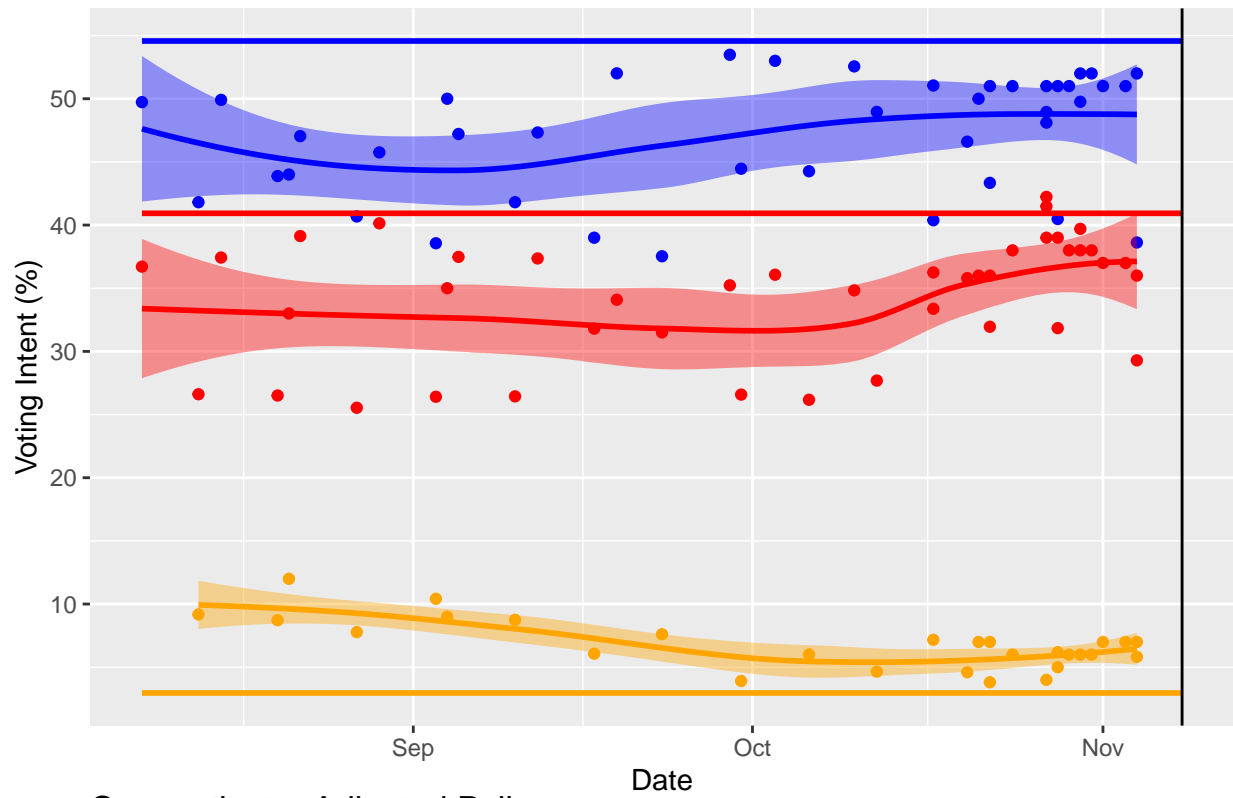
Colorado – Raw Polls



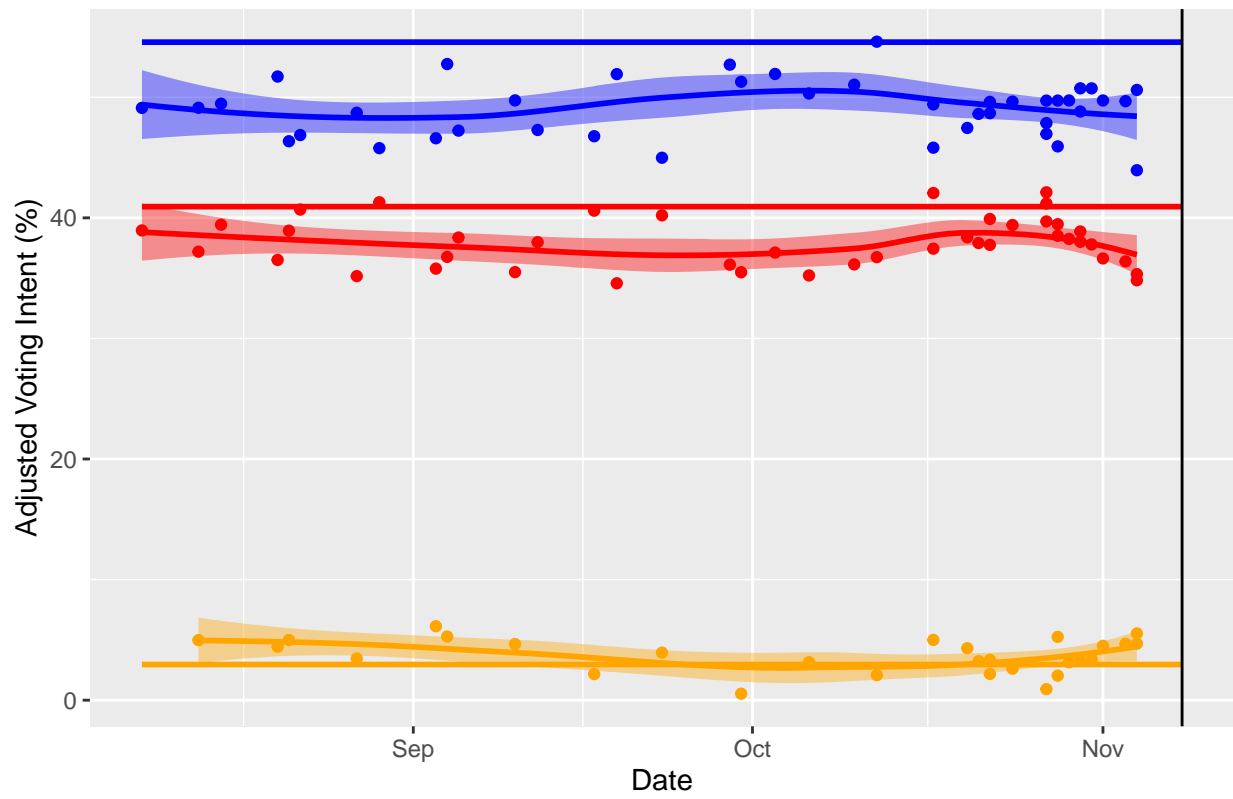
Colorado – Adjusted Polls



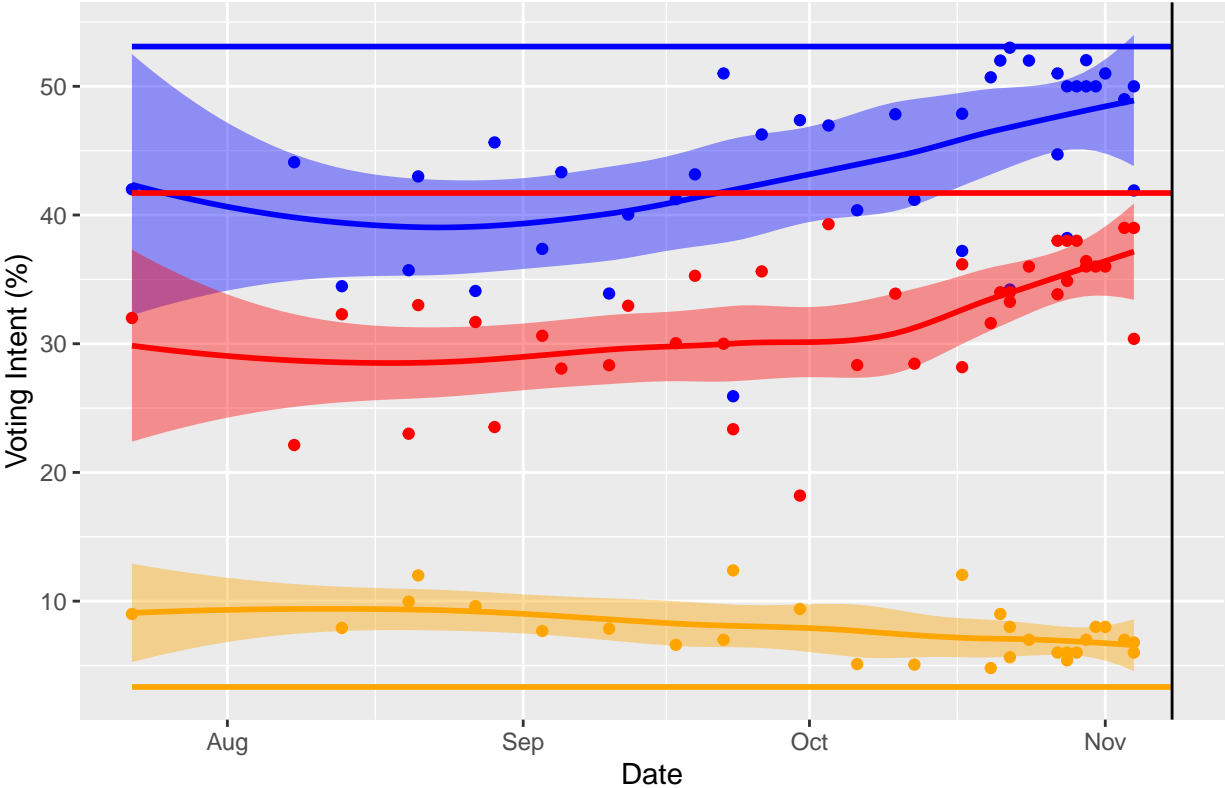
Connecticut – Raw Polls



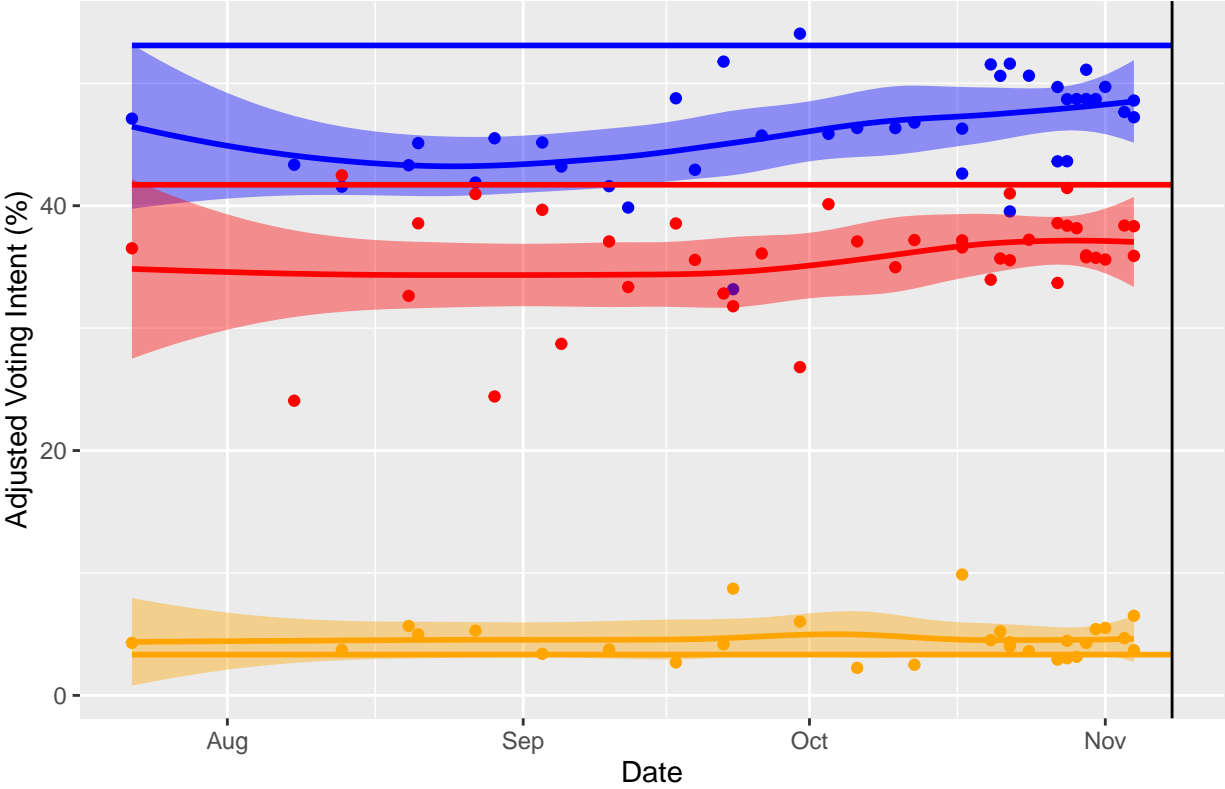
Connecticut – Adjusted Polls



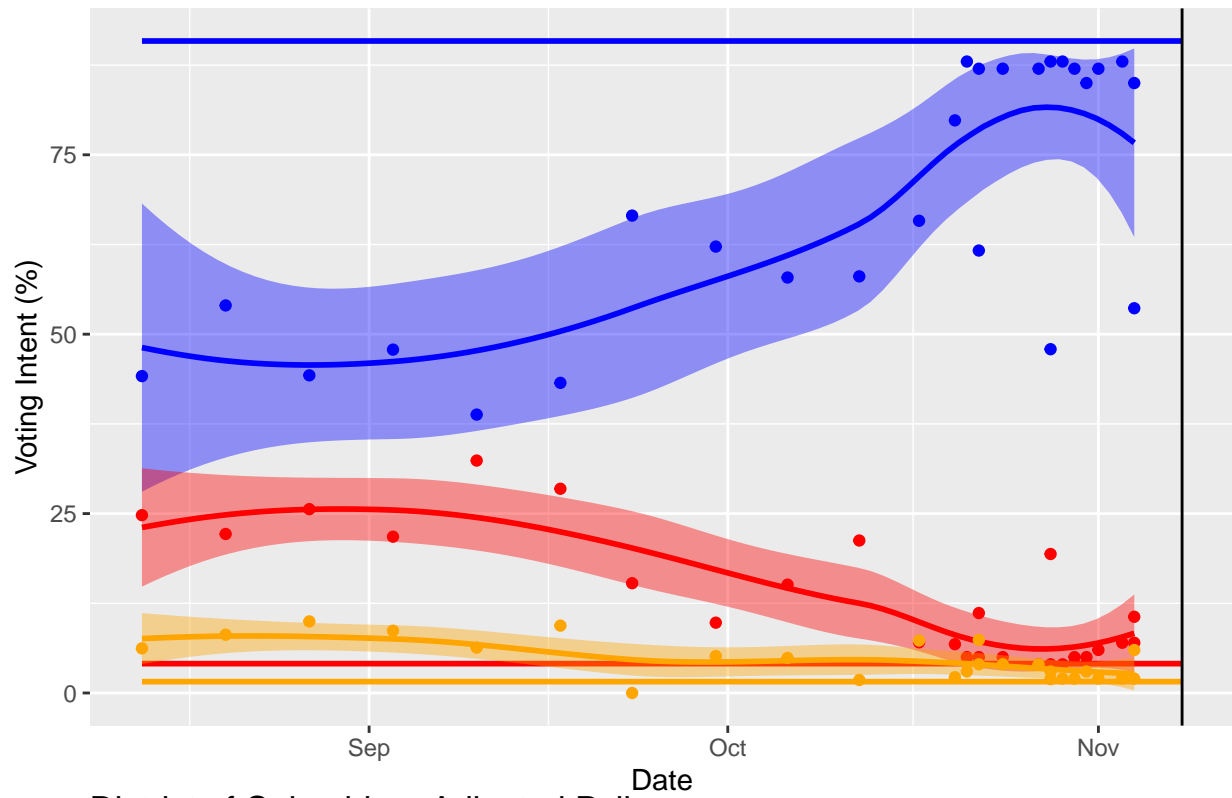
Delaware – Raw Polls



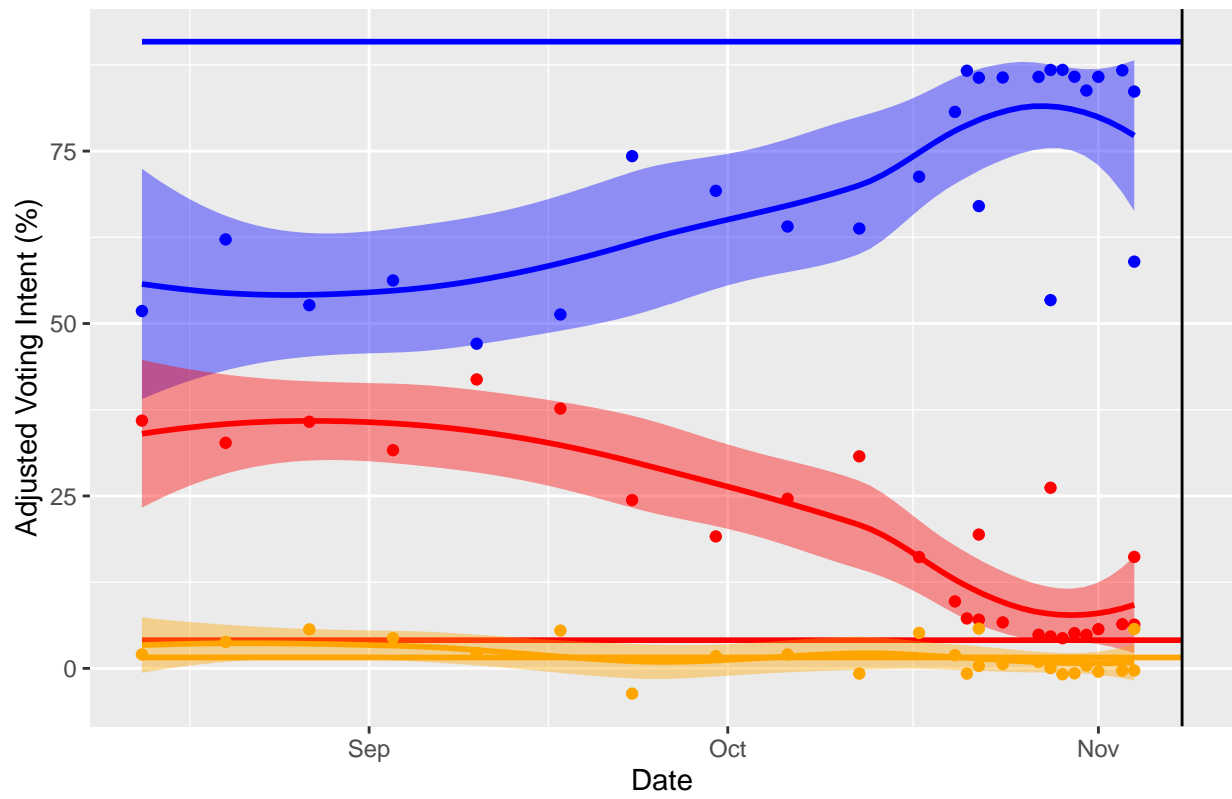
Delaware – Adjusted Polls



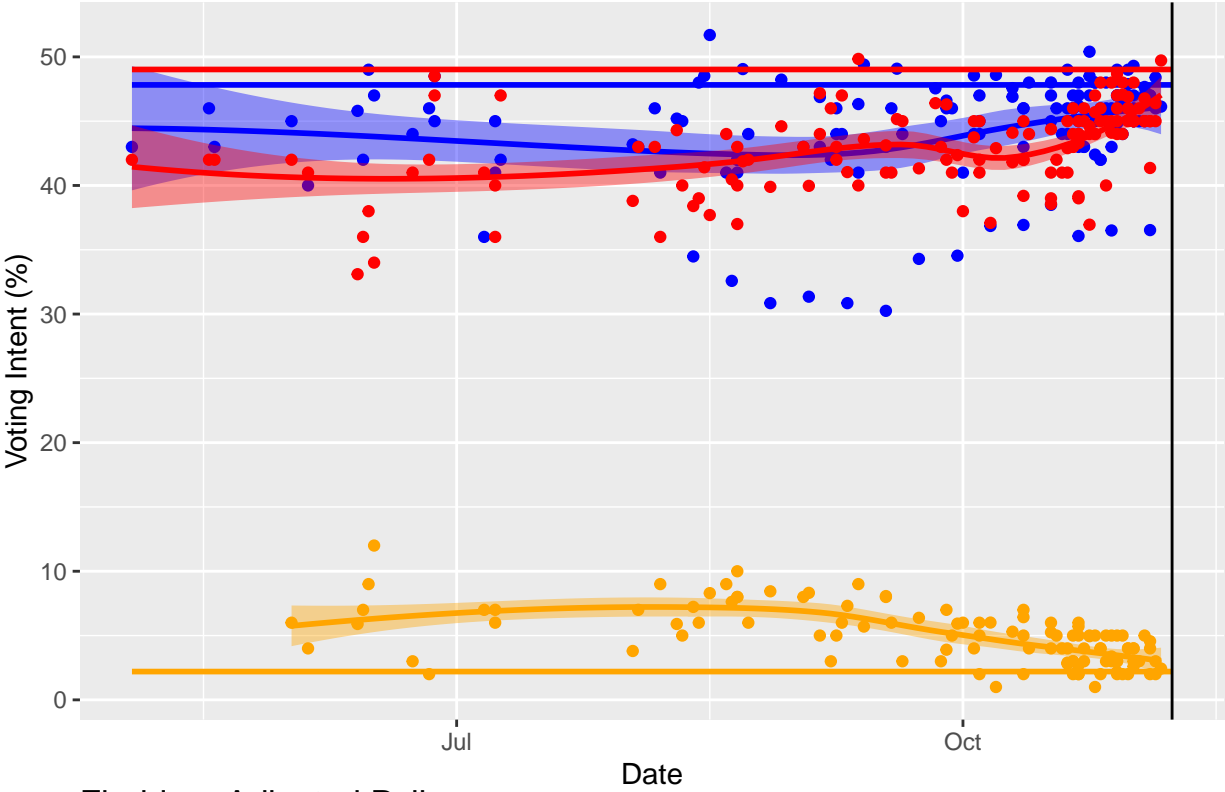
District of Columbia – Raw Polls



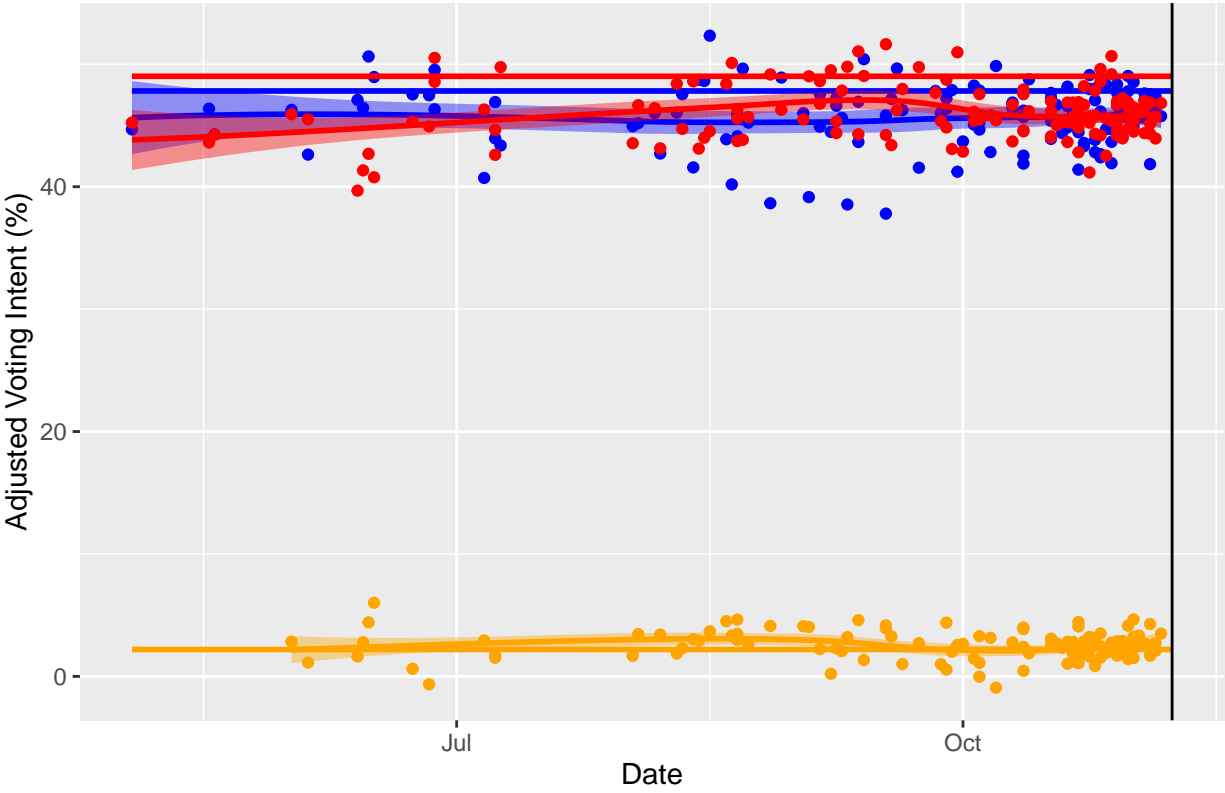
District of Columbia – Adjusted Polls



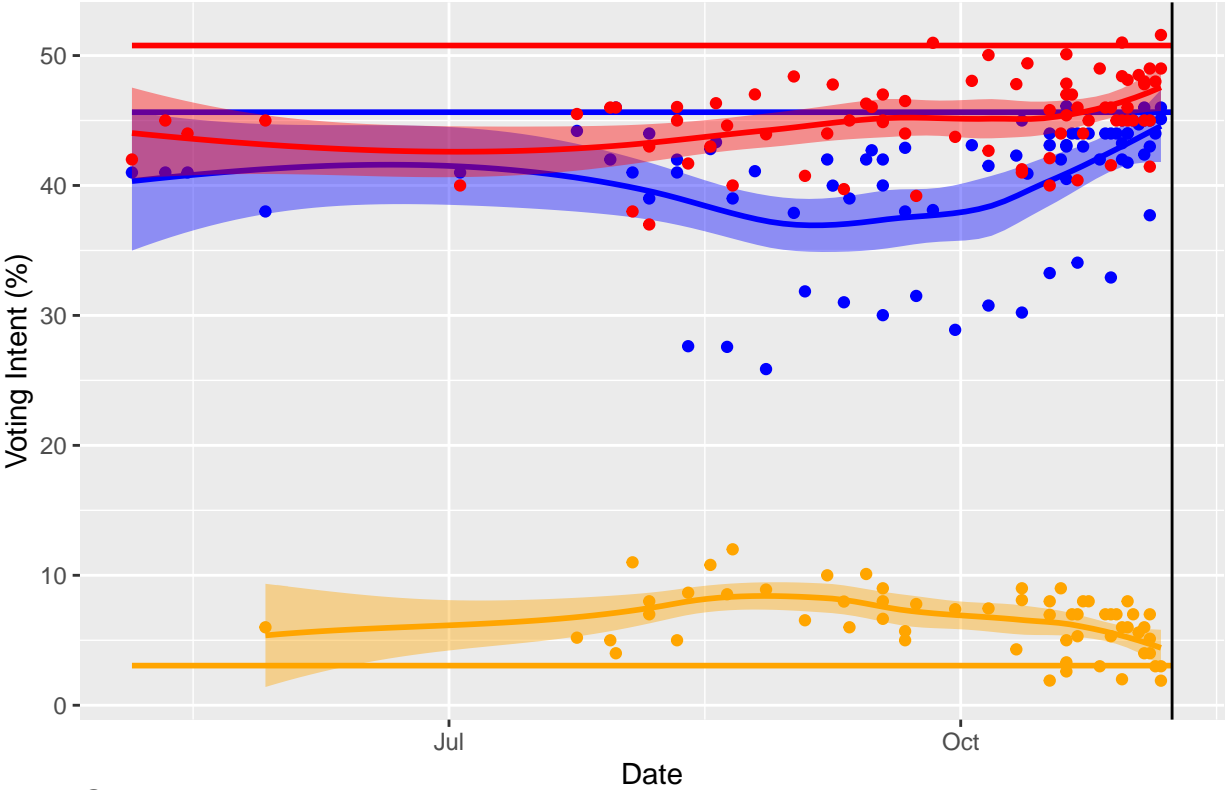
Florida – Raw Polls



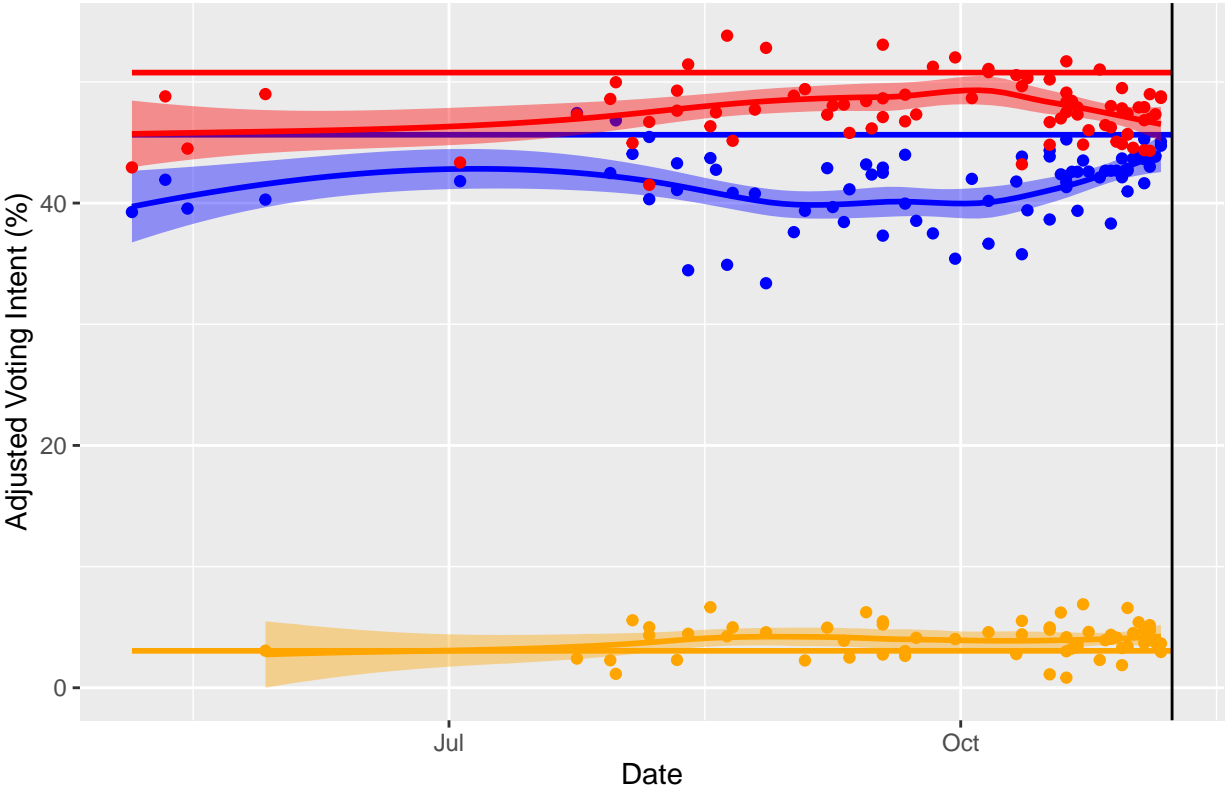
Florida – Adjusted Polls



Georgia – Raw Polls

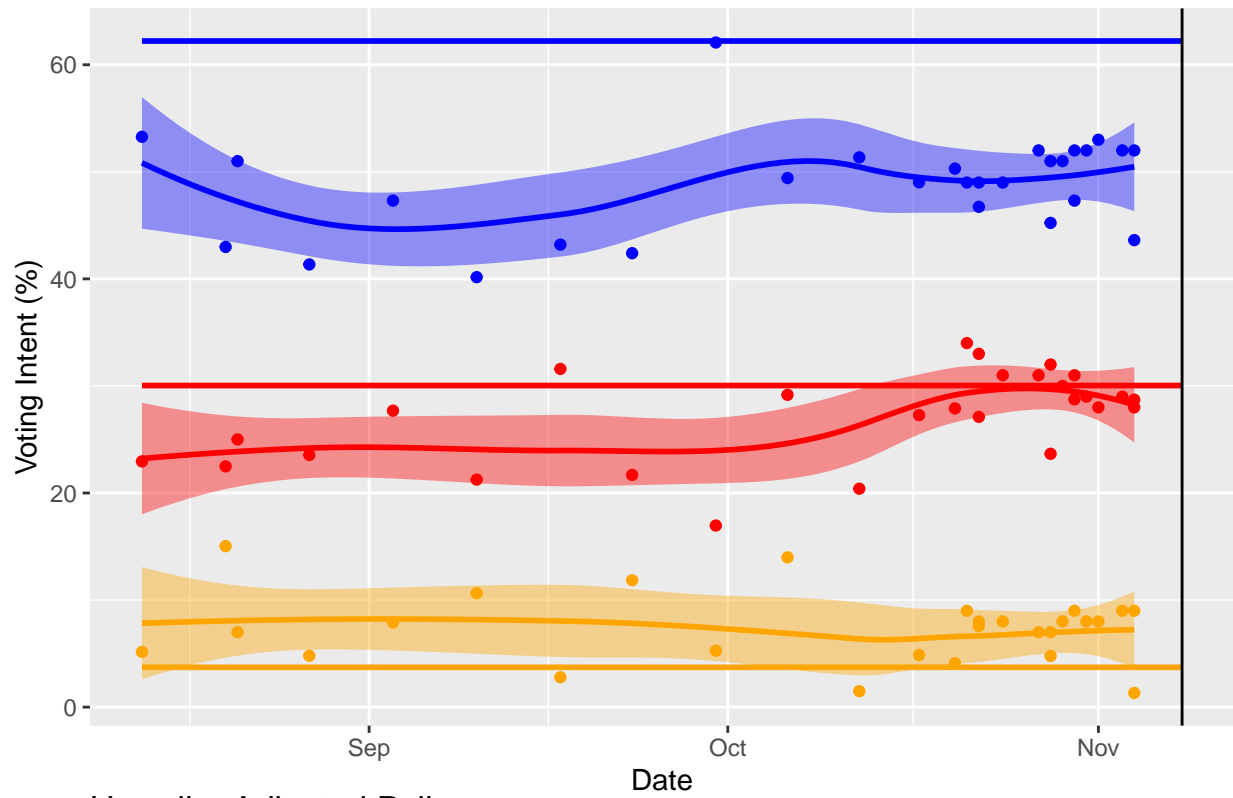


Georgia – Adjusted Polls

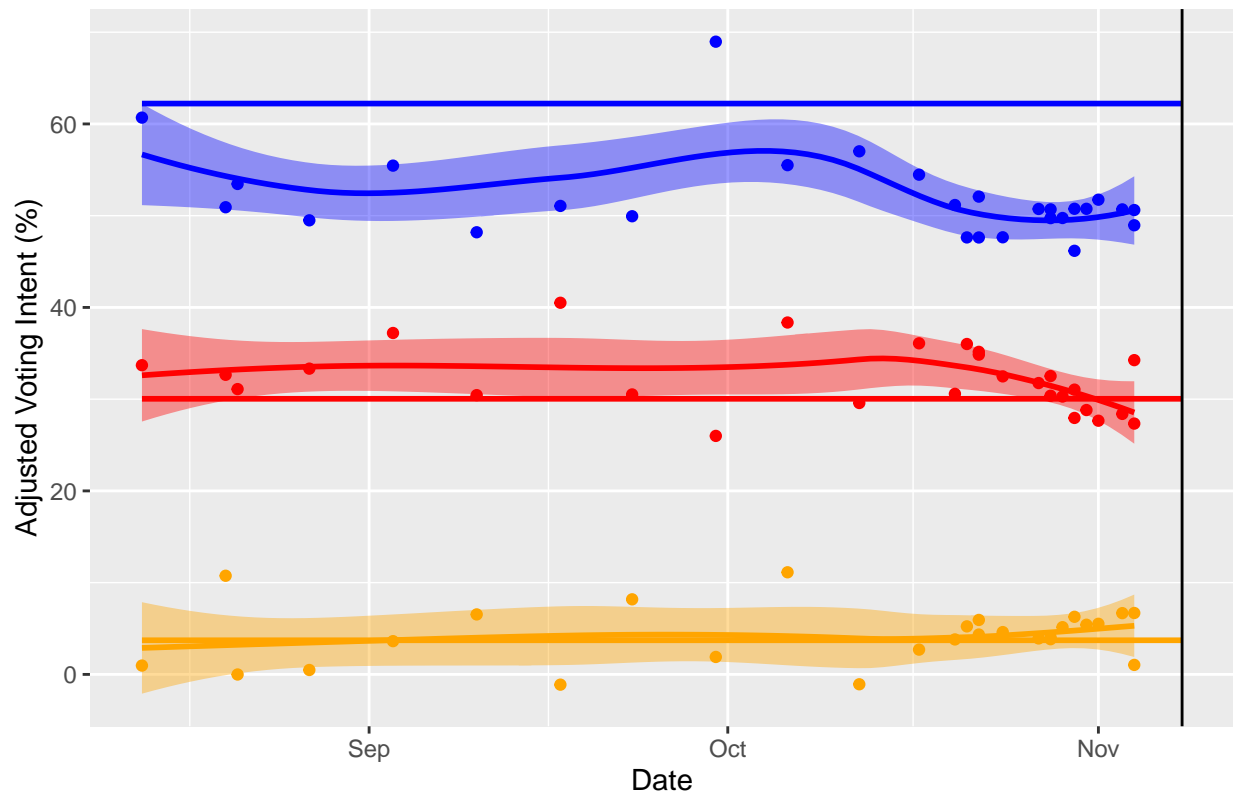




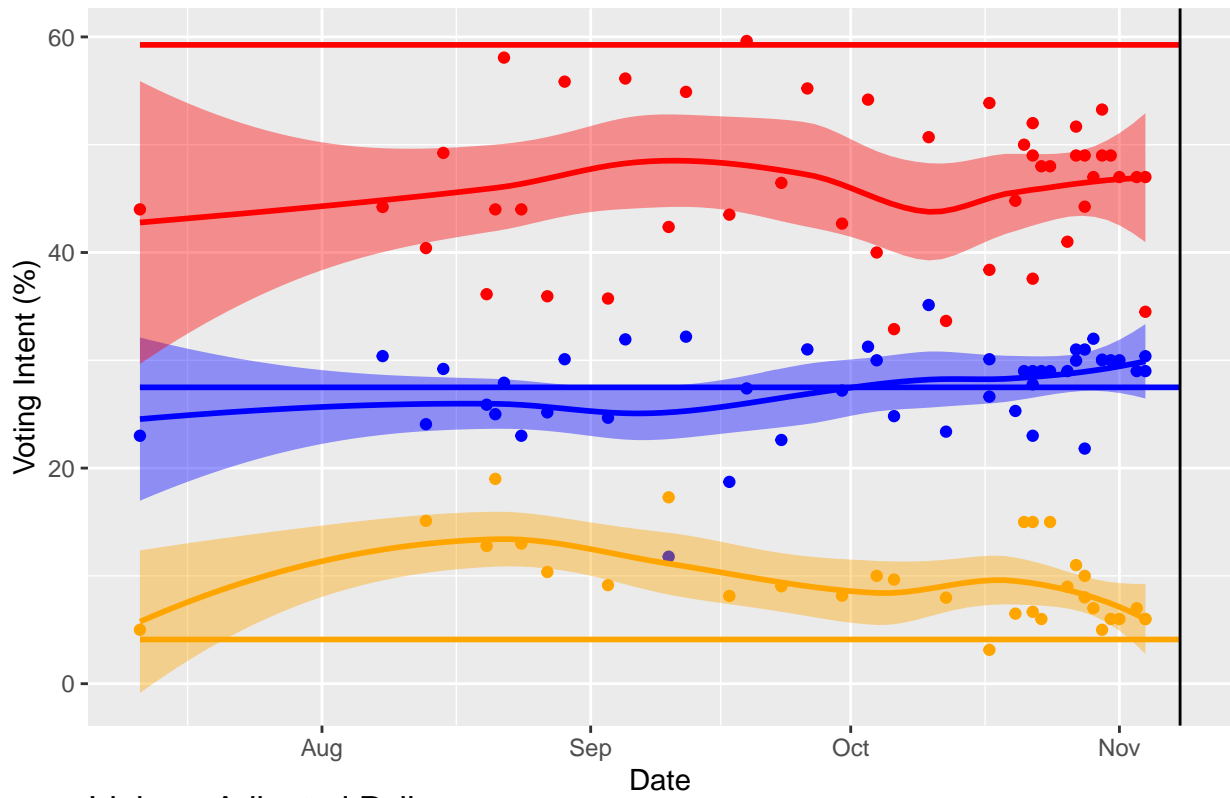
Hawaii – Raw Polls



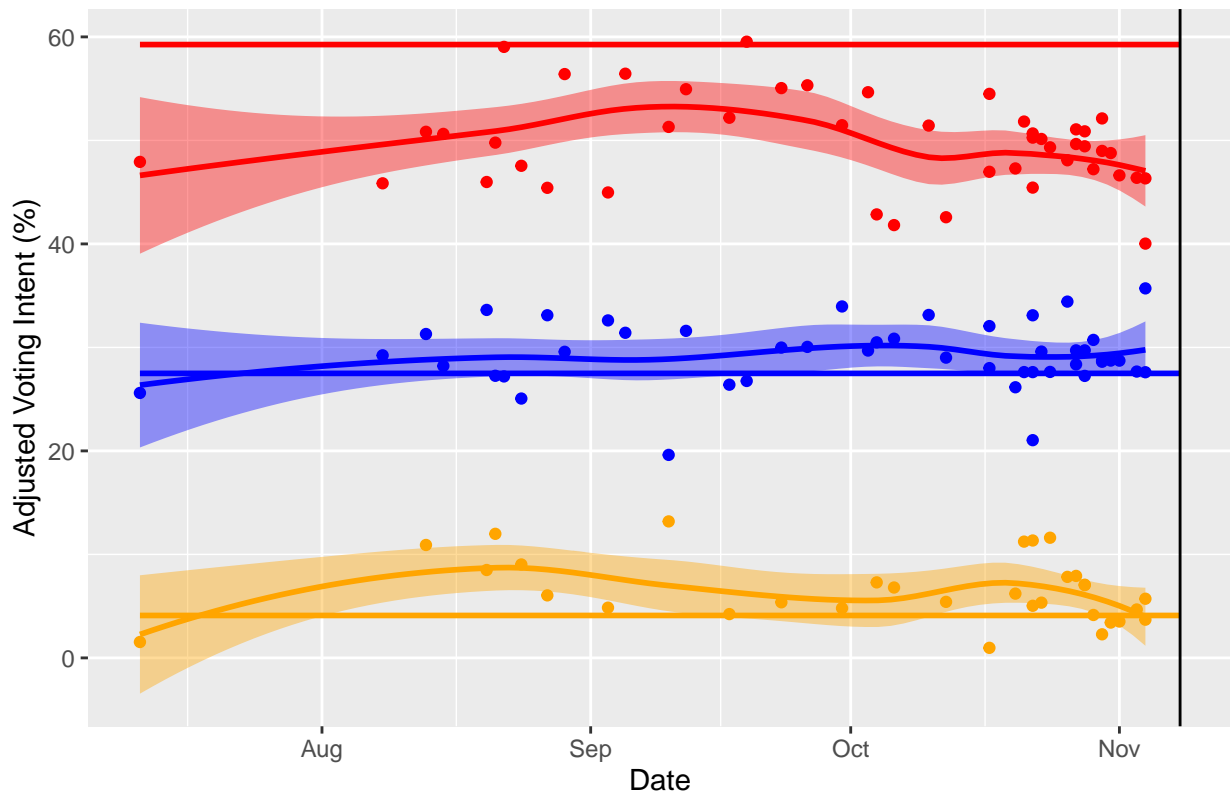
Hawaii – Adjusted Polls



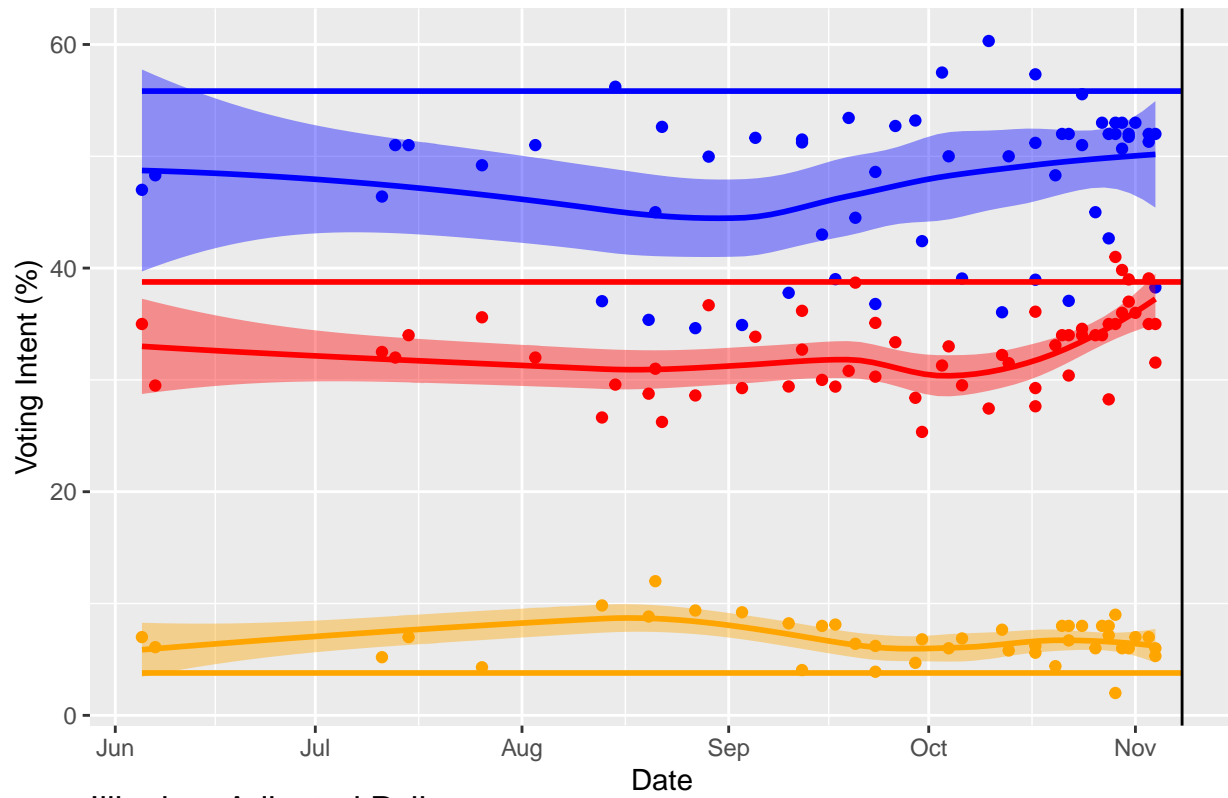
Idaho – Raw Polls



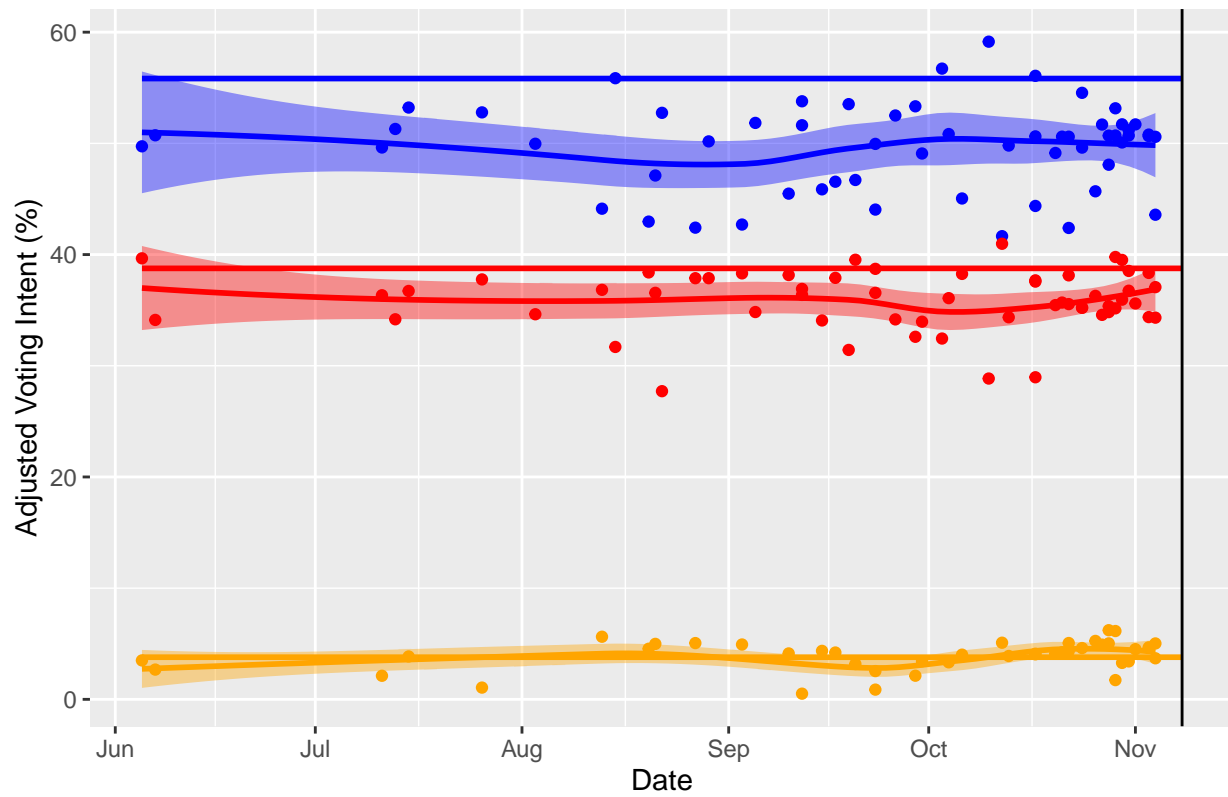
Idaho – Adjusted Polls



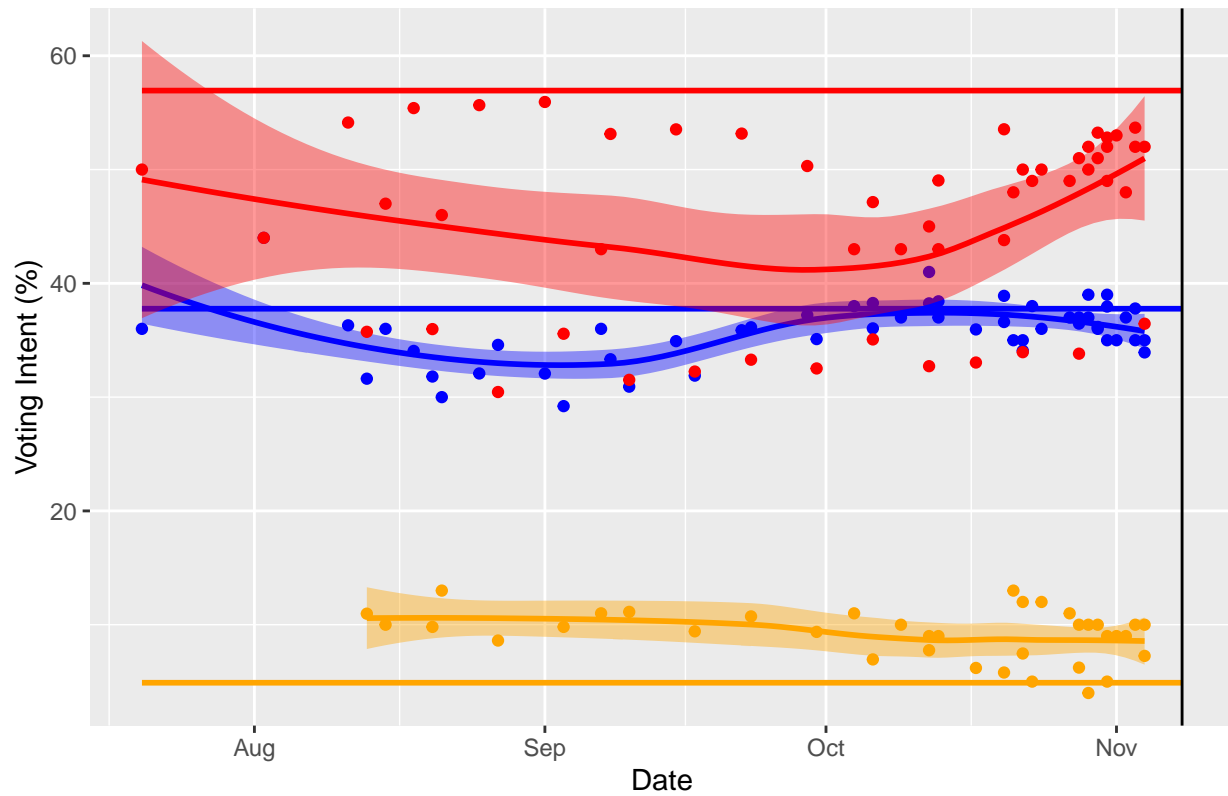
Illinois – Raw Polls



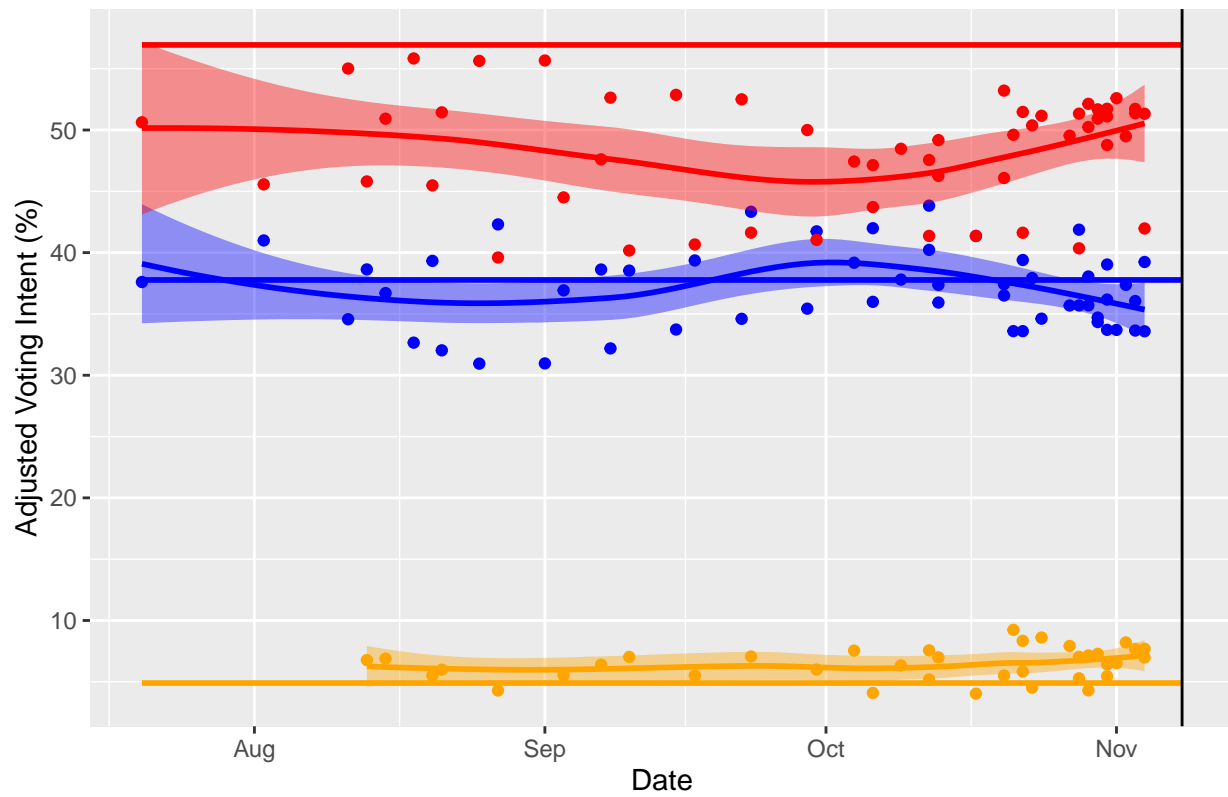
Illinois – Adjusted Polls



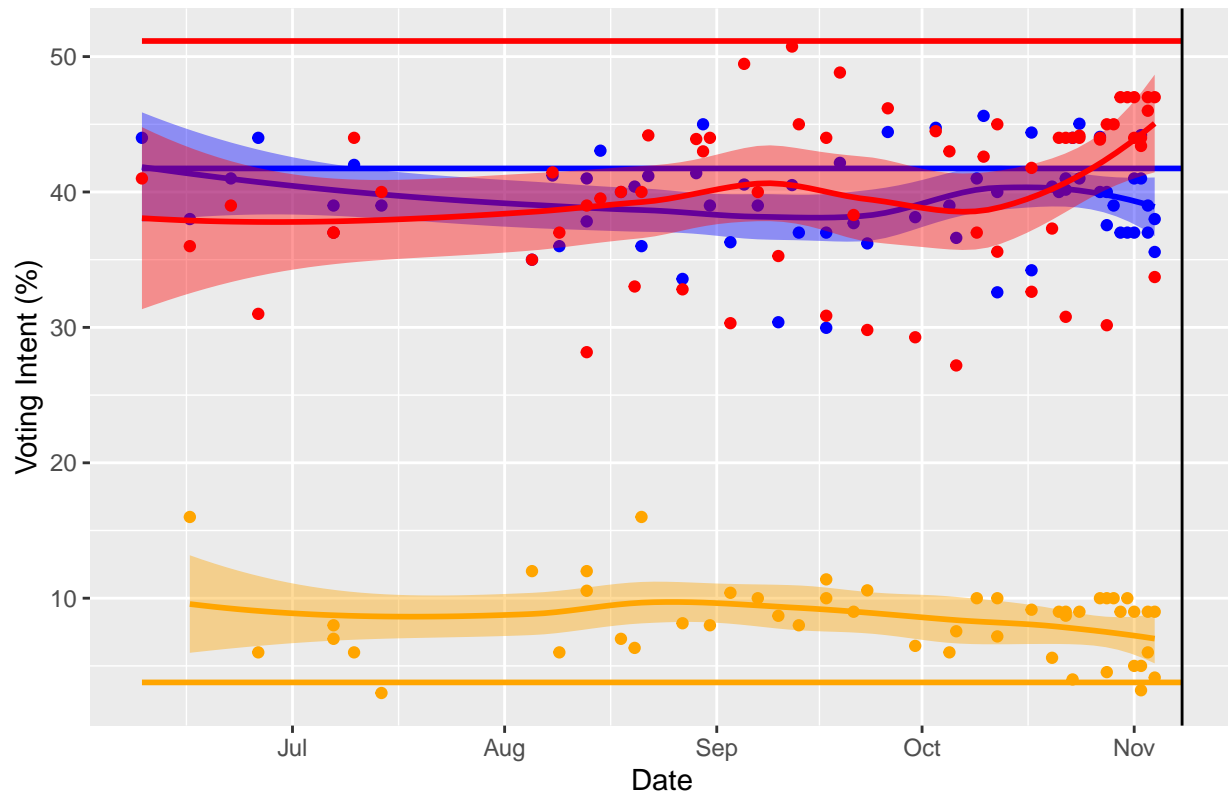
Indiana – Raw Polls



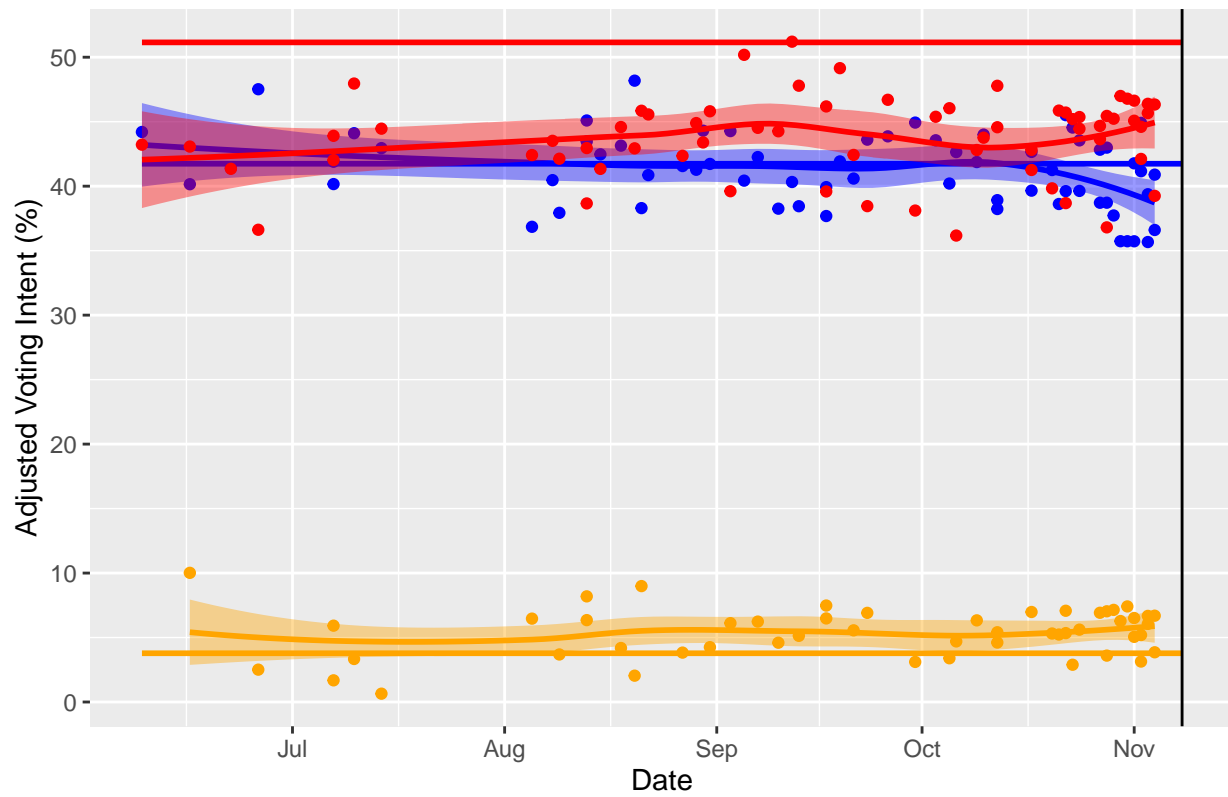
Indiana – Adjusted Polls



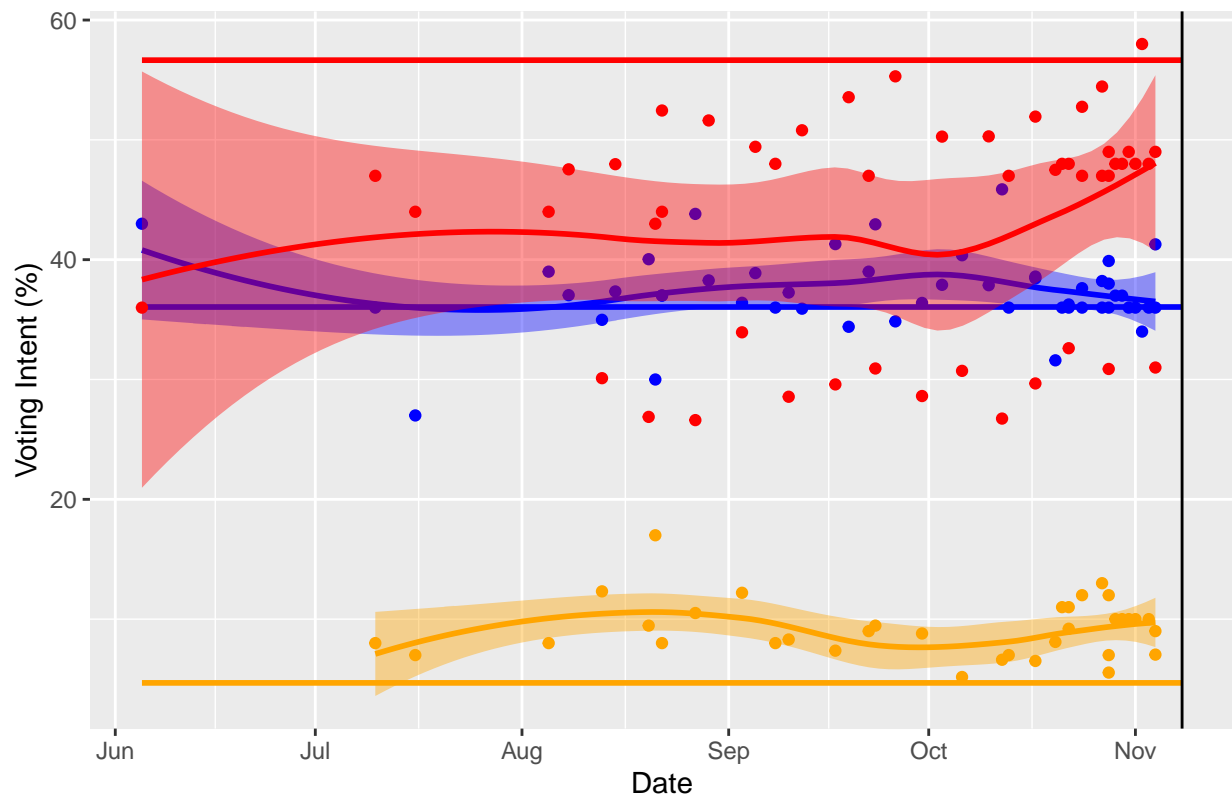
Iowa – Raw Polls



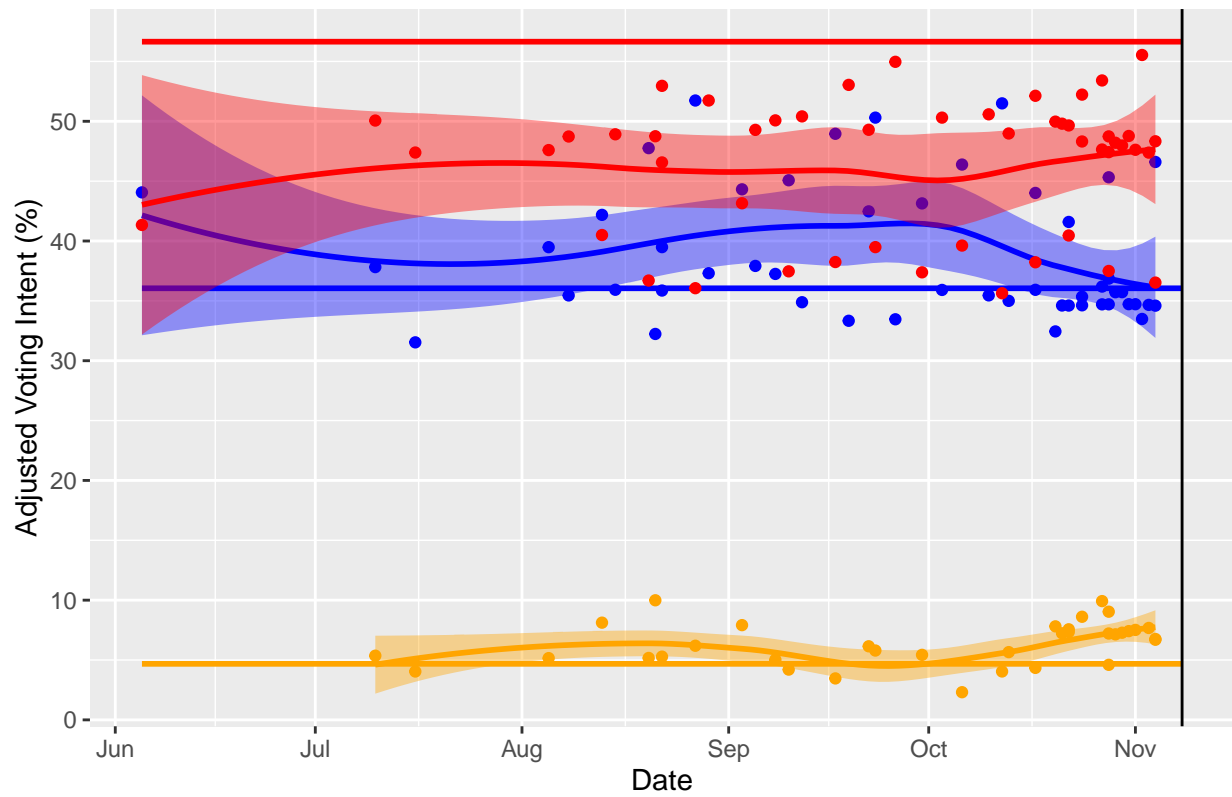
Iowa – Adjusted Polls



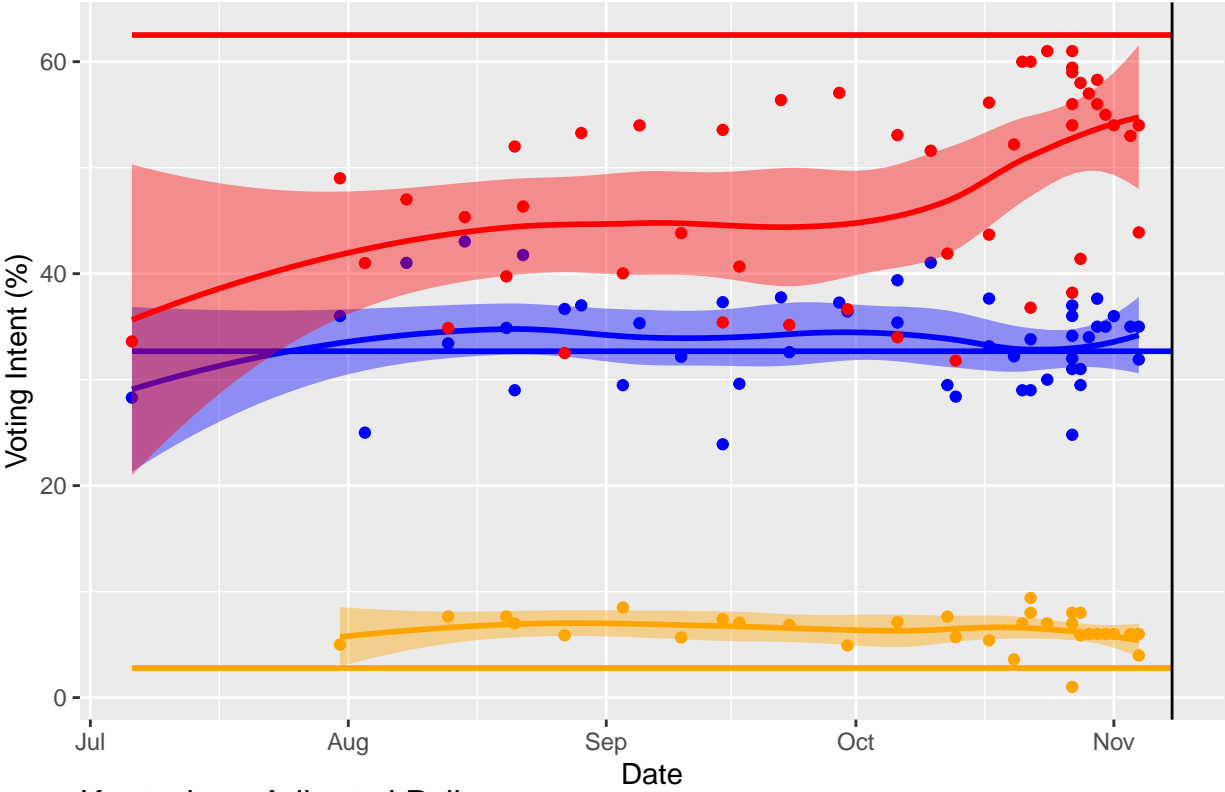
Kansas – Raw Polls



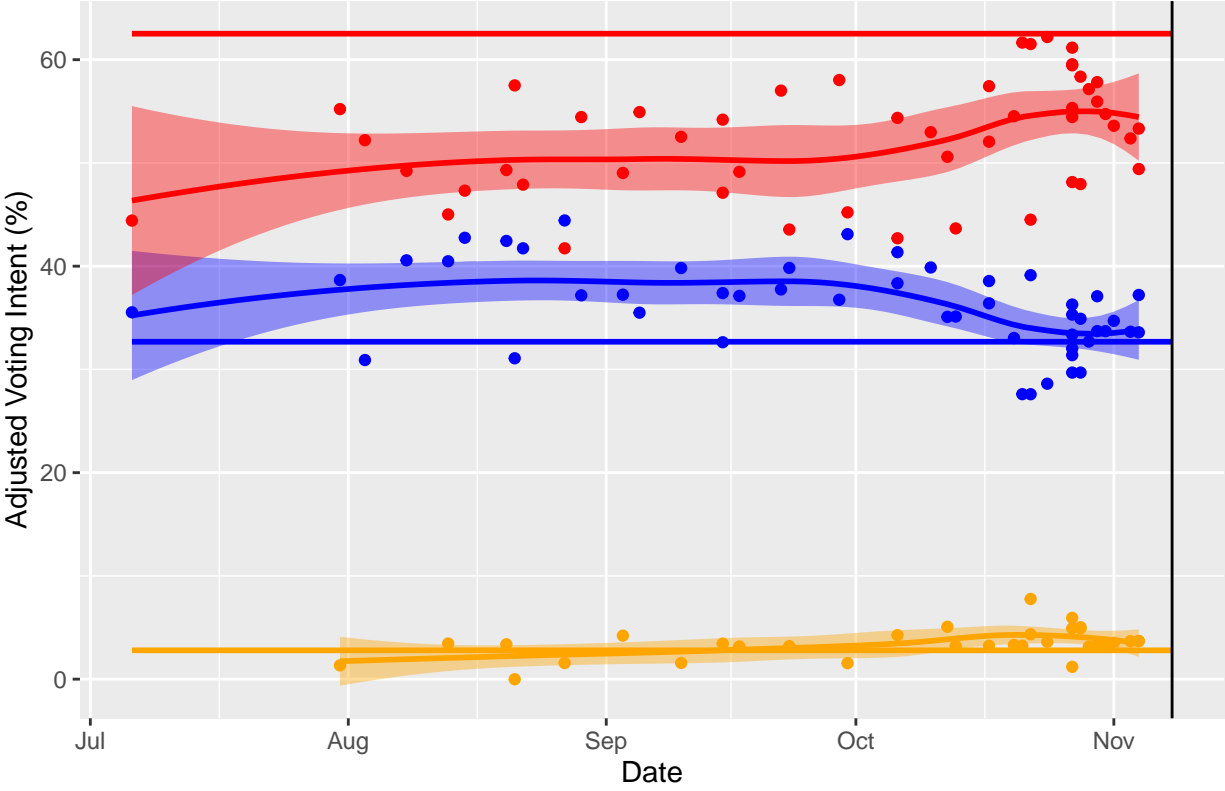
Kansas – Adjusted Polls



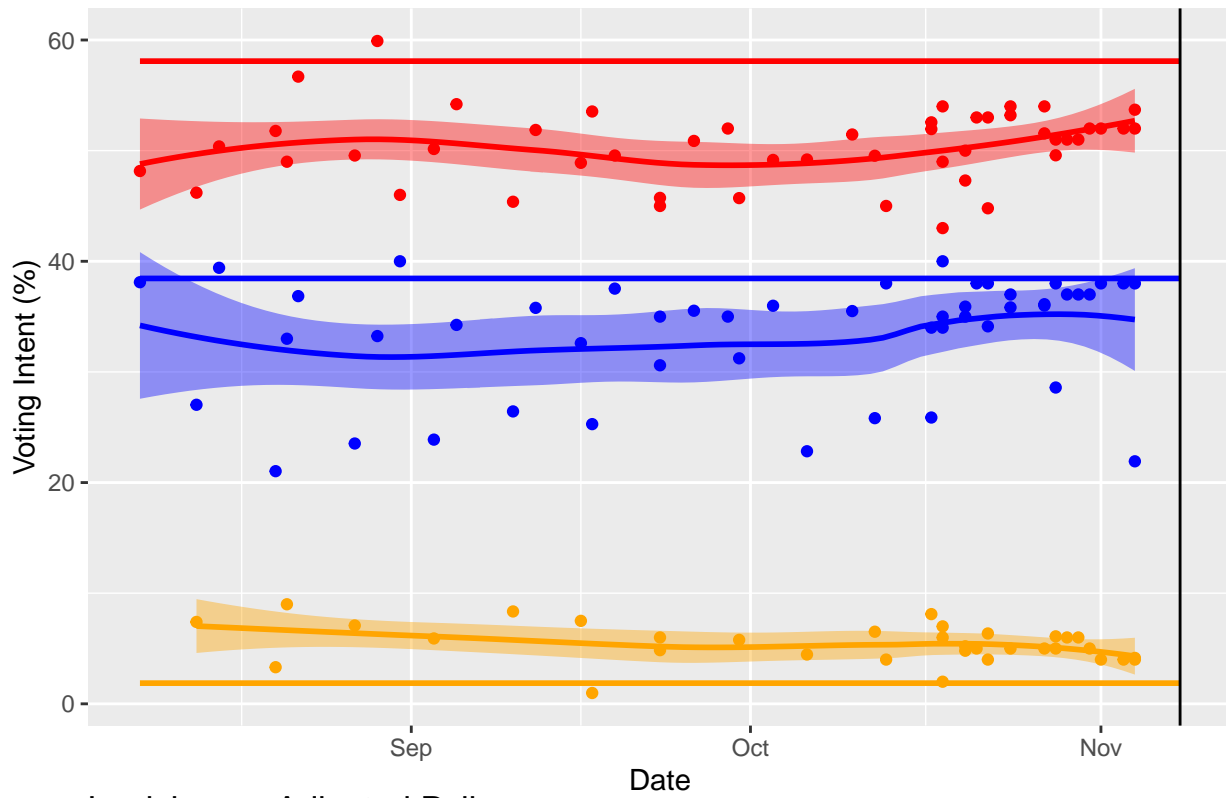
Kentucky – Raw Polls



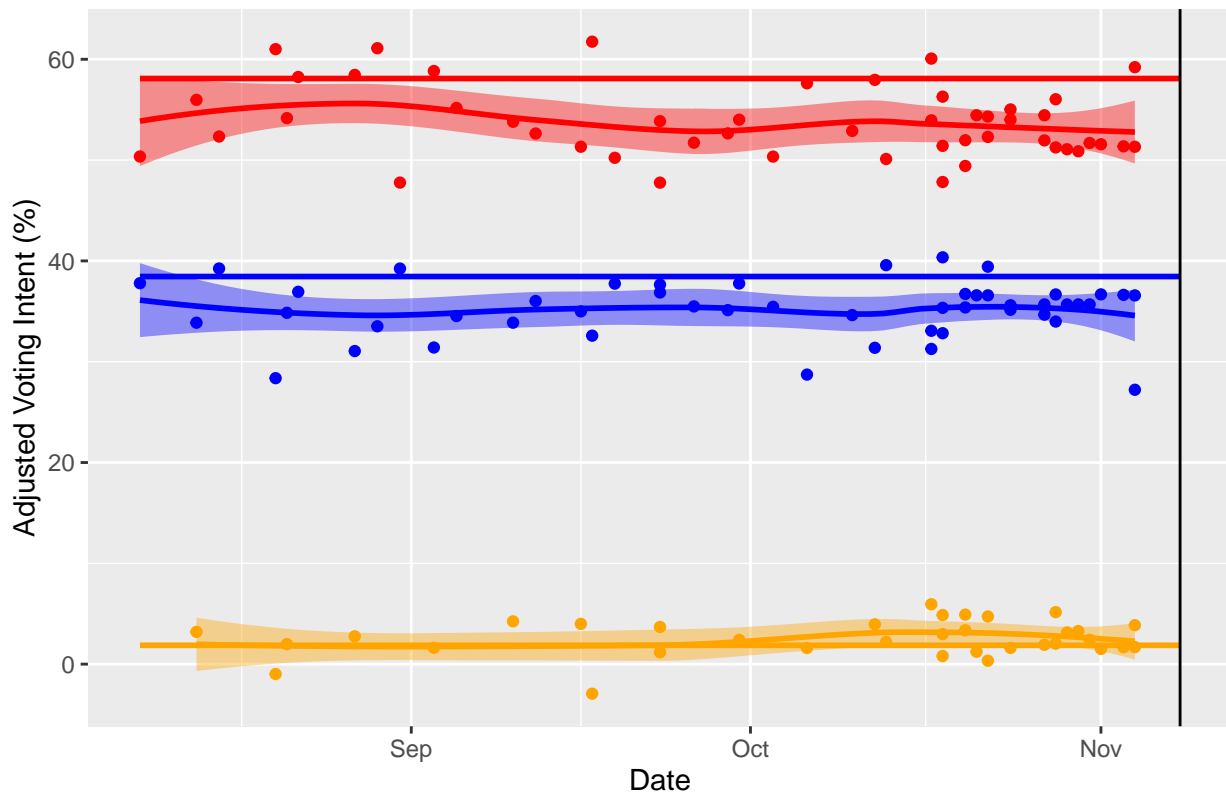
Kentucky – Adjusted Polls



Louisiana – Raw Polls

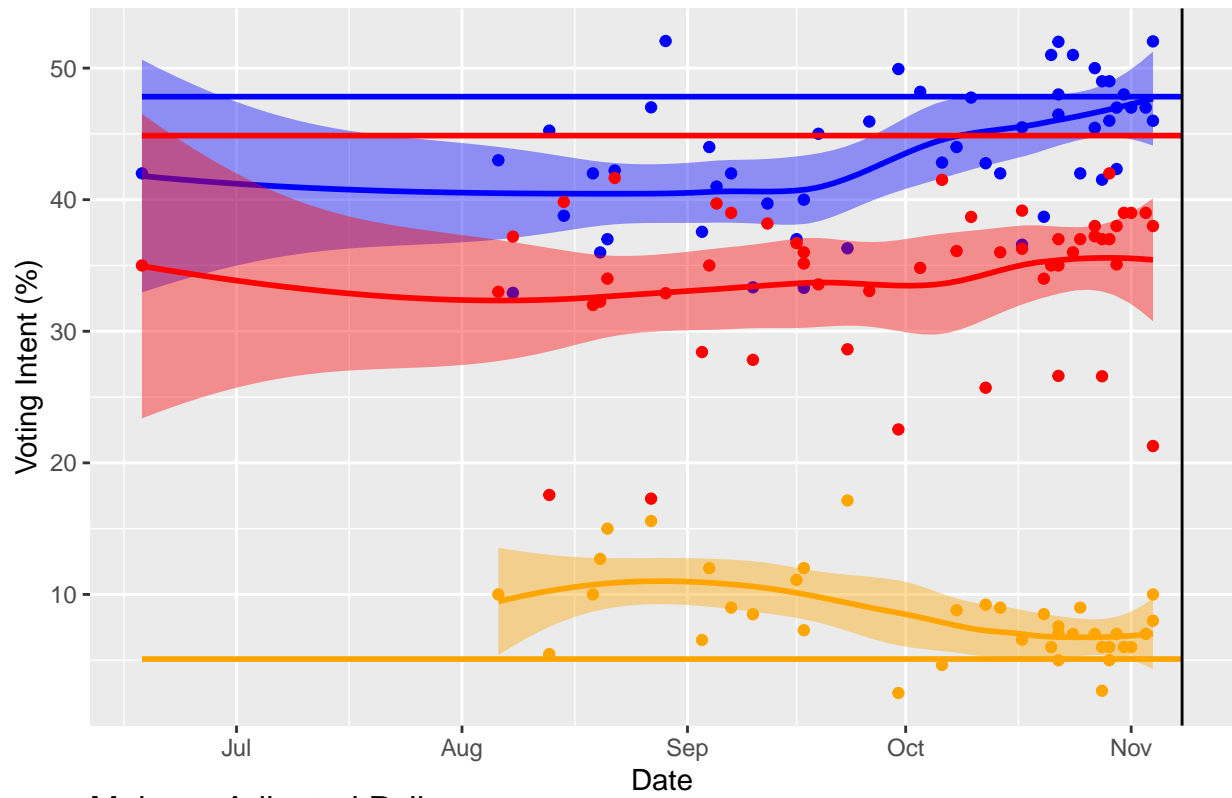


Louisiana – Adjusted Polls

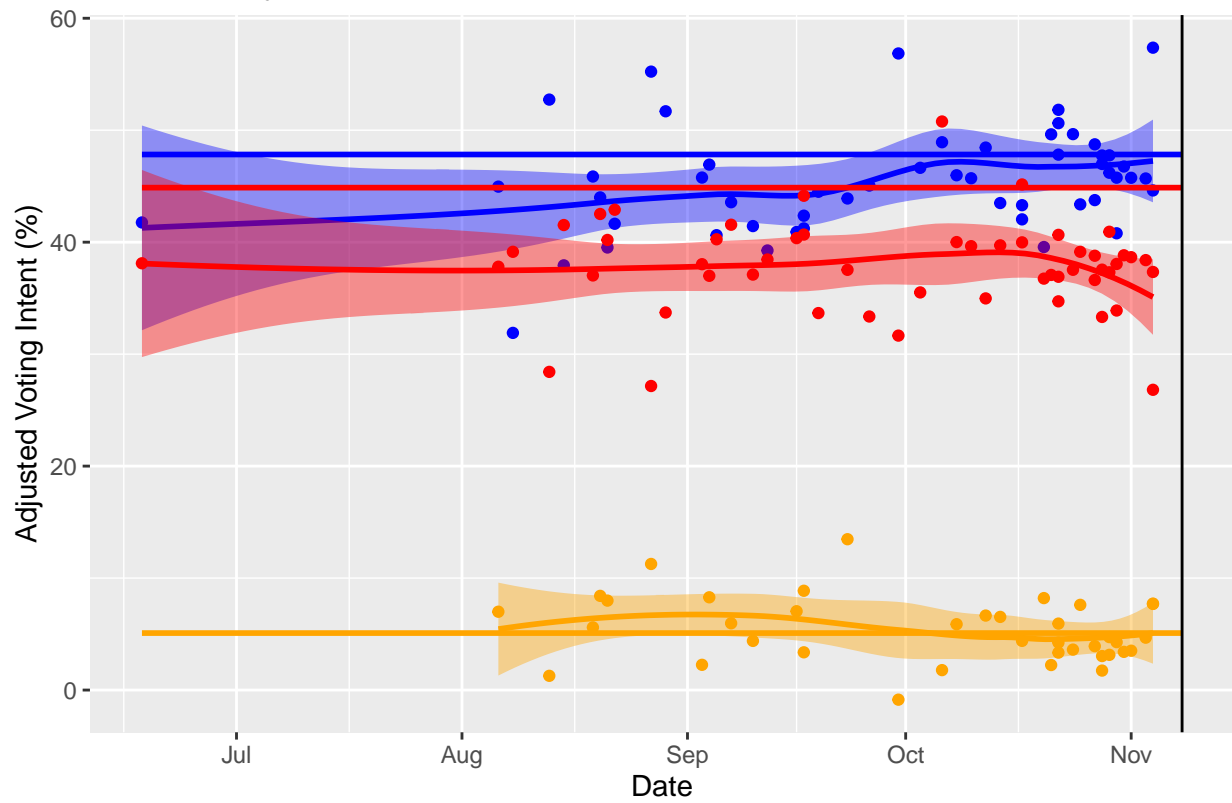




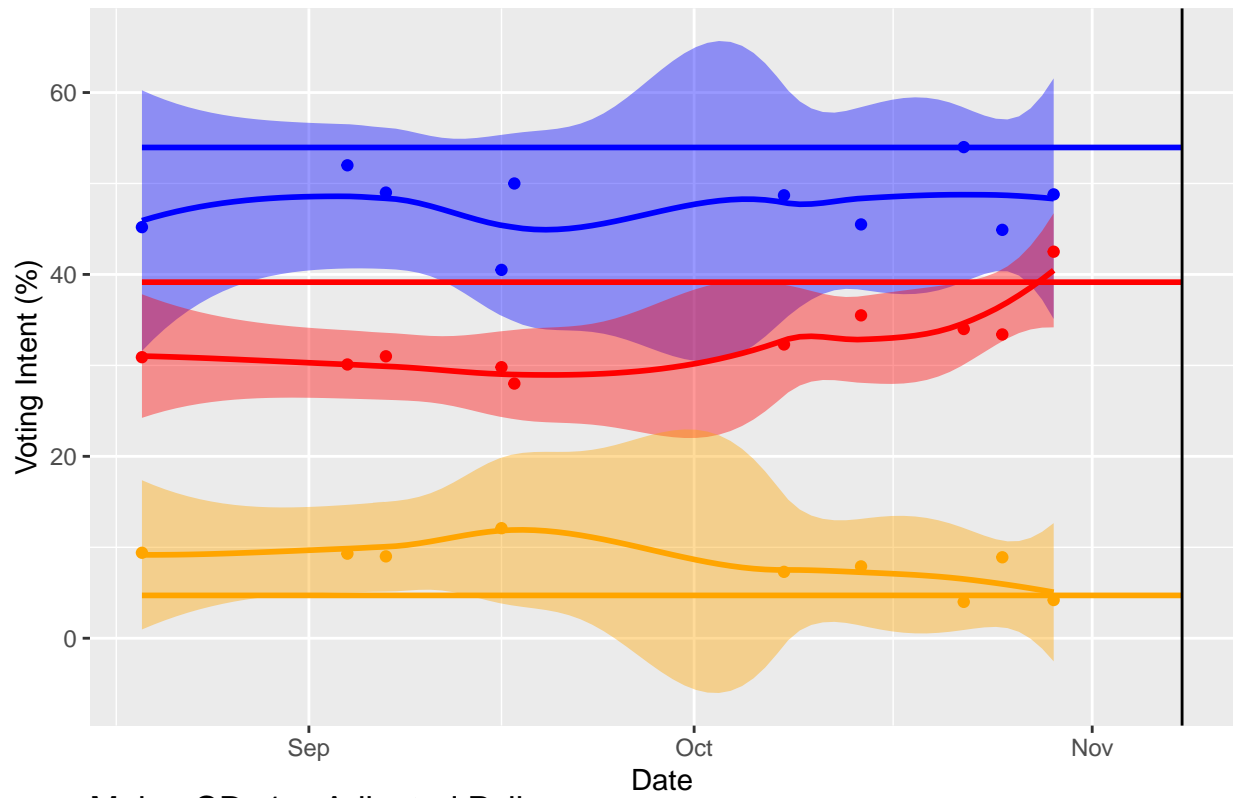
Maine – Raw Polls



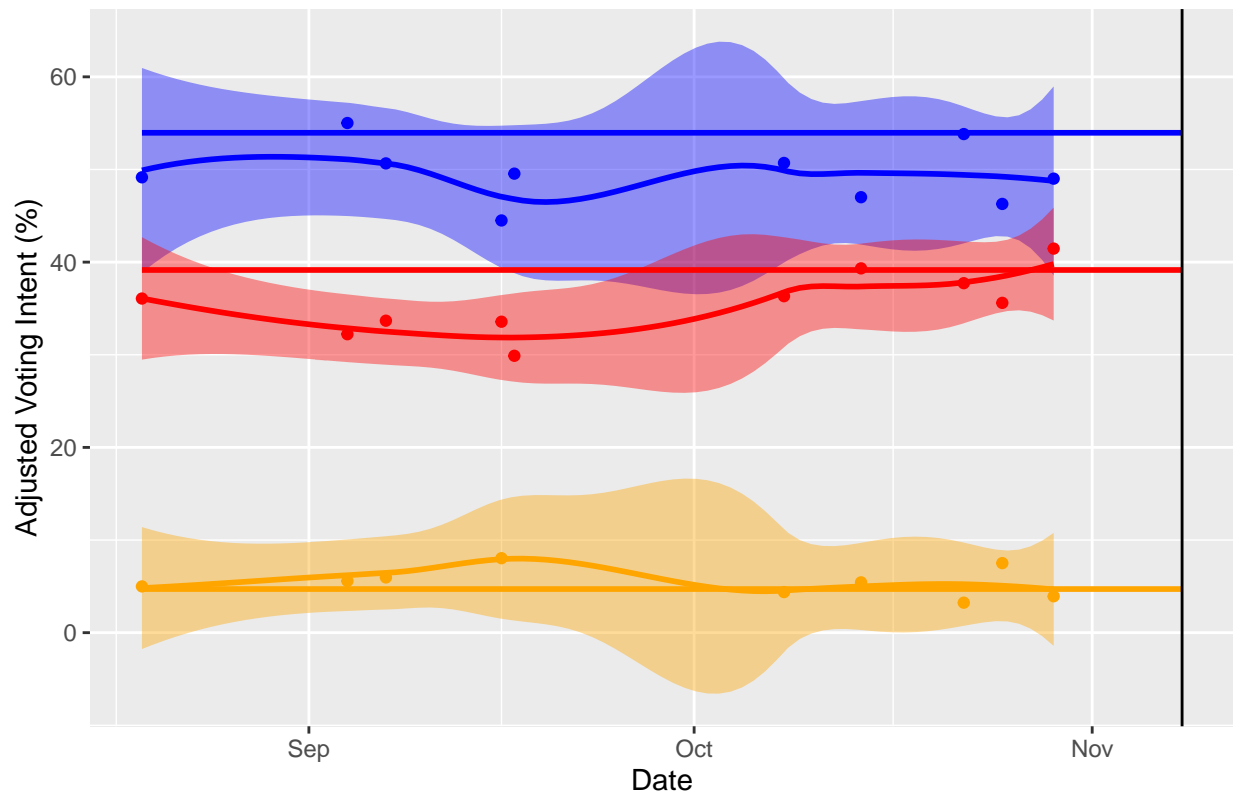
Maine – Adjusted Polls



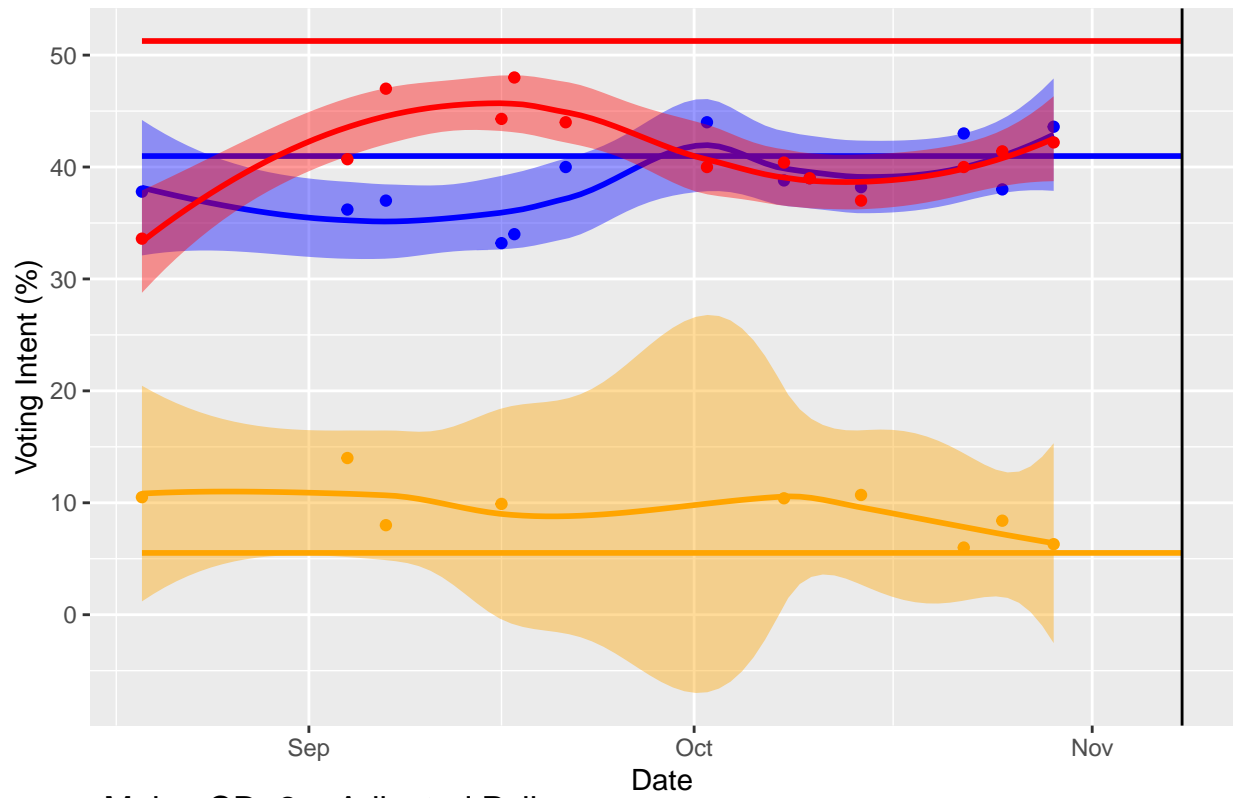
Maine CD-1 – Raw Polls



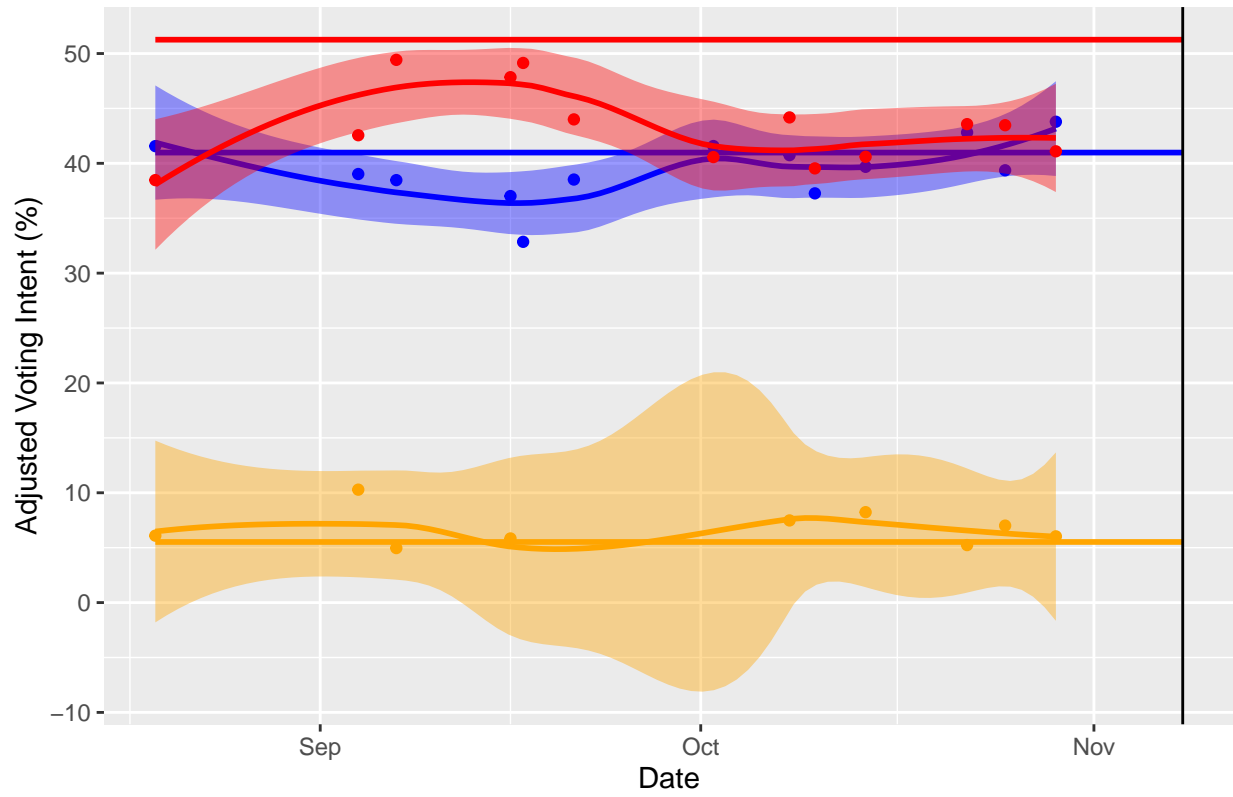
Maine CD-1 – Adjusted Polls



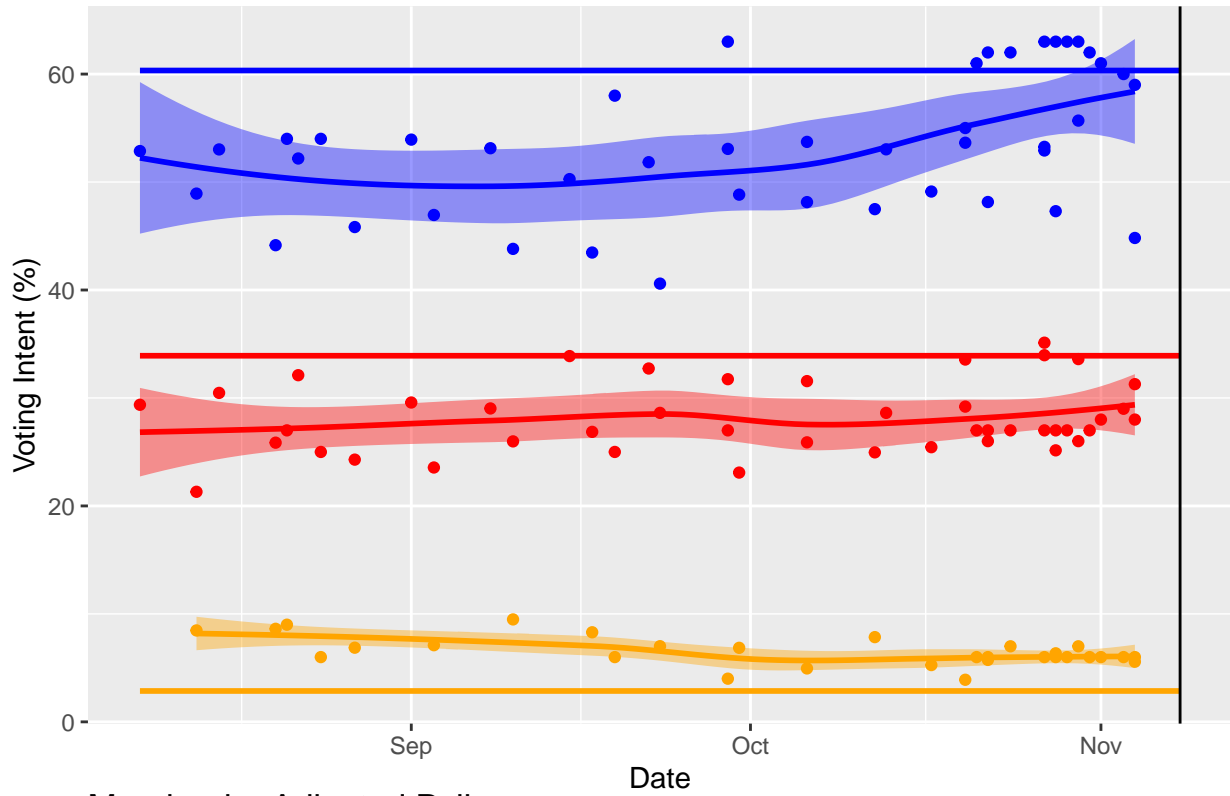
Maine CD-2 – Raw Polls



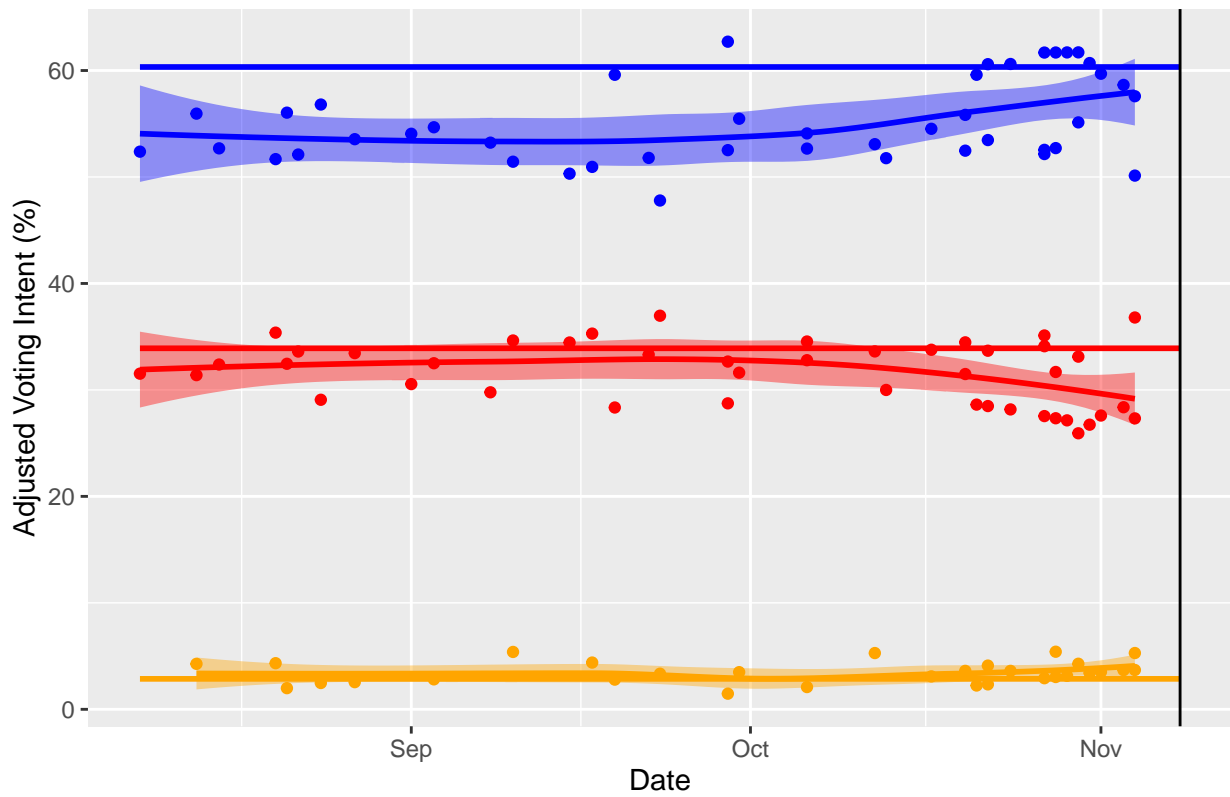
Maine CD-2 – Adjusted Polls



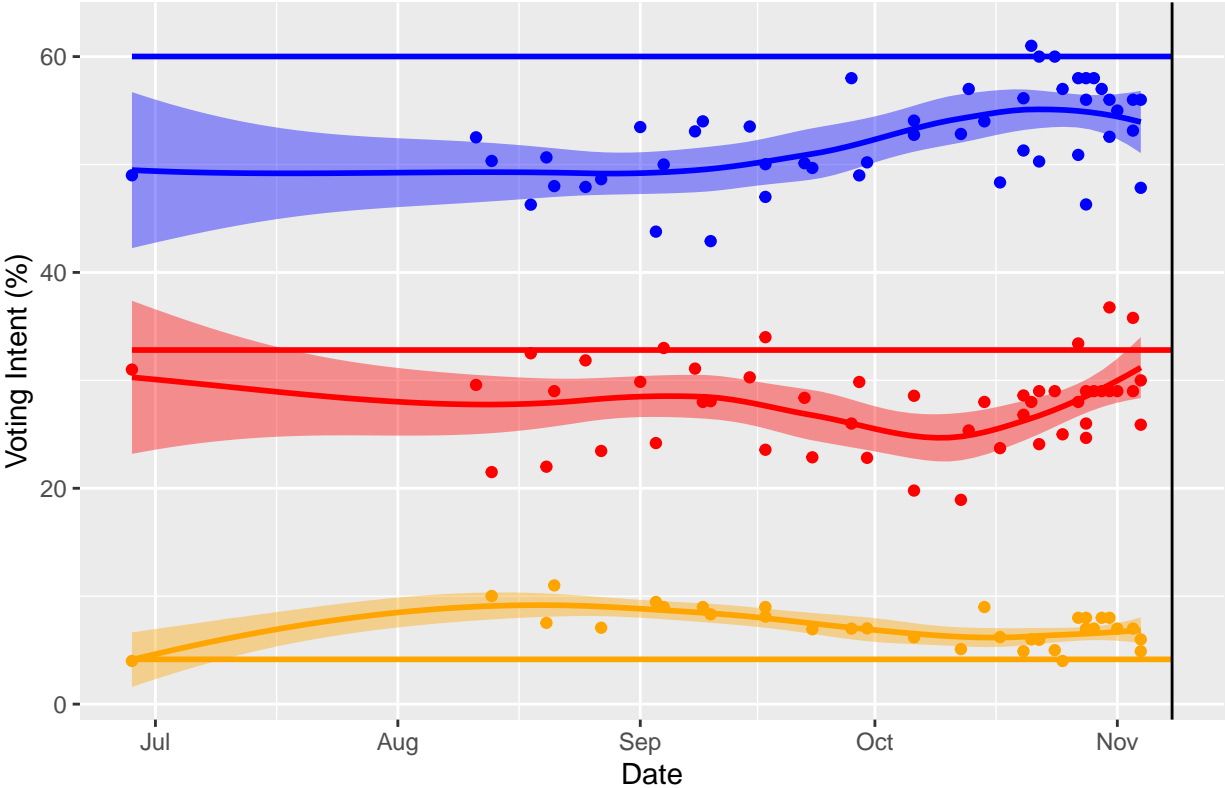
Maryland – Raw Polls



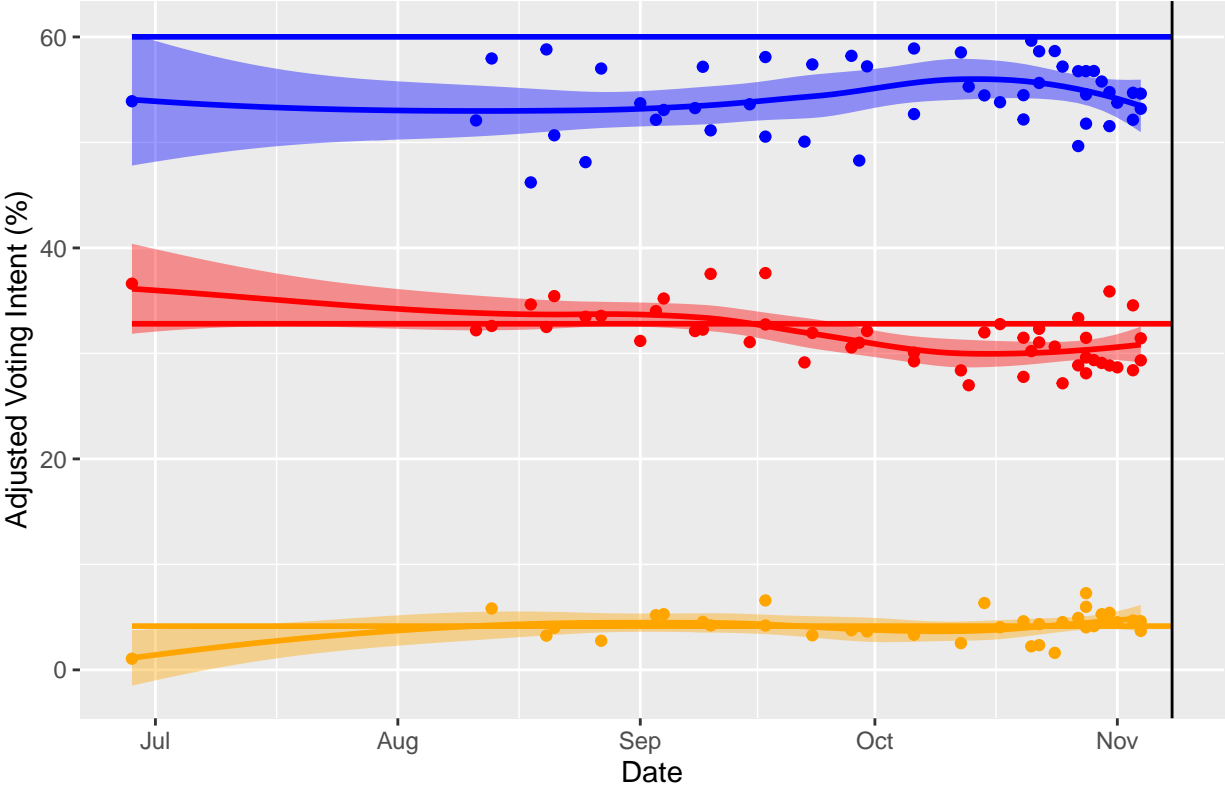
Maryland – Adjusted Polls



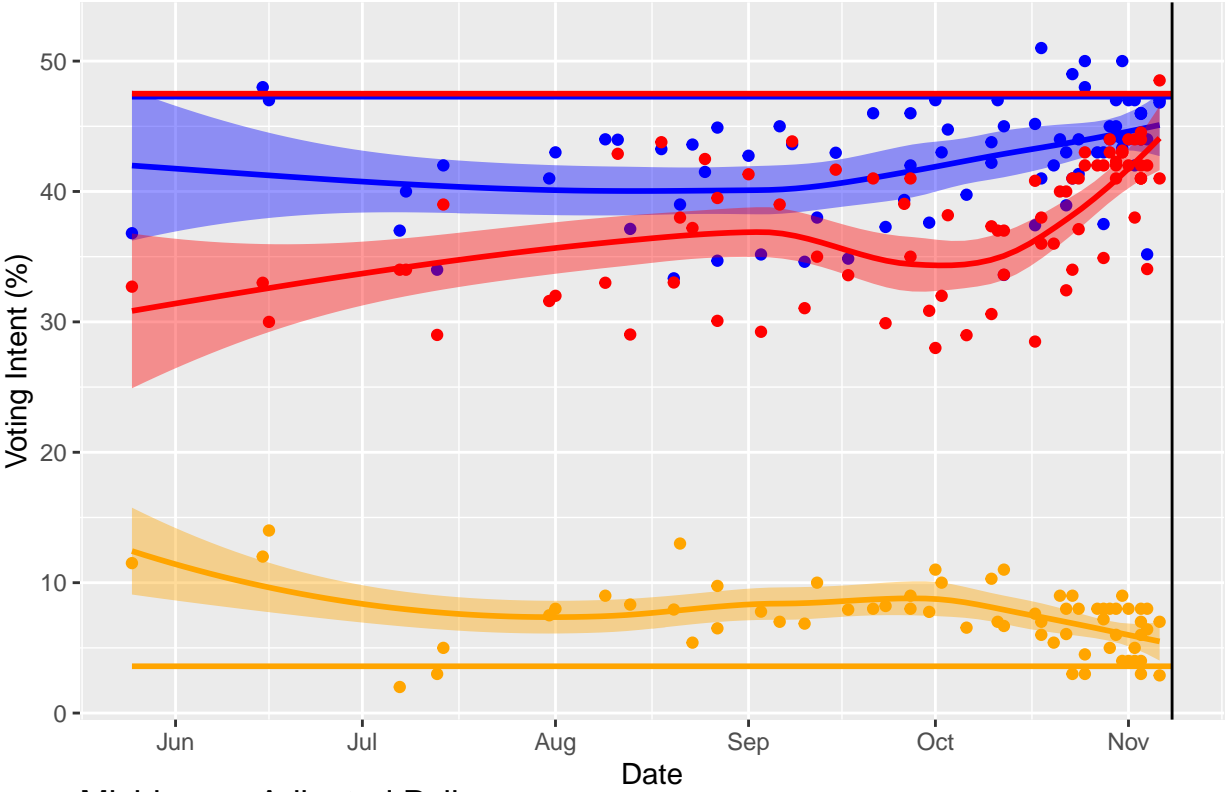
Massachusetts – Raw Polls



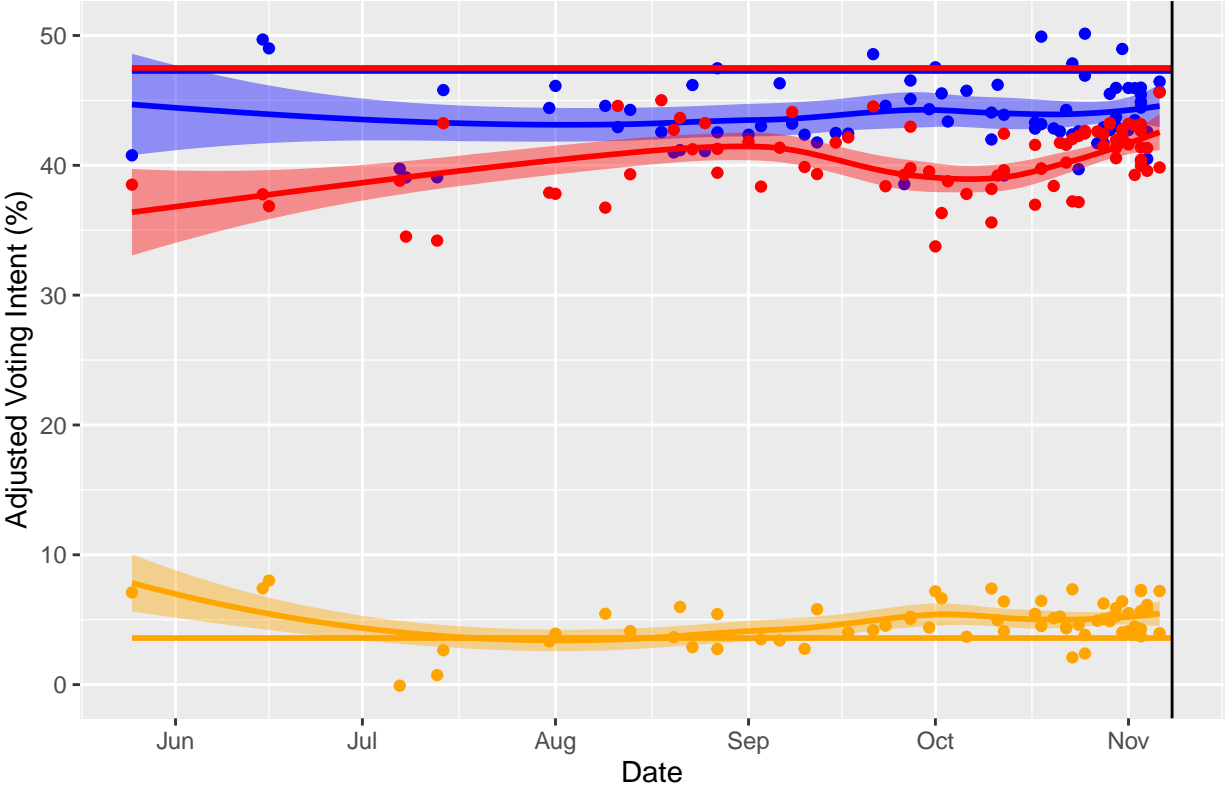
Massachusetts – Adjusted Polls



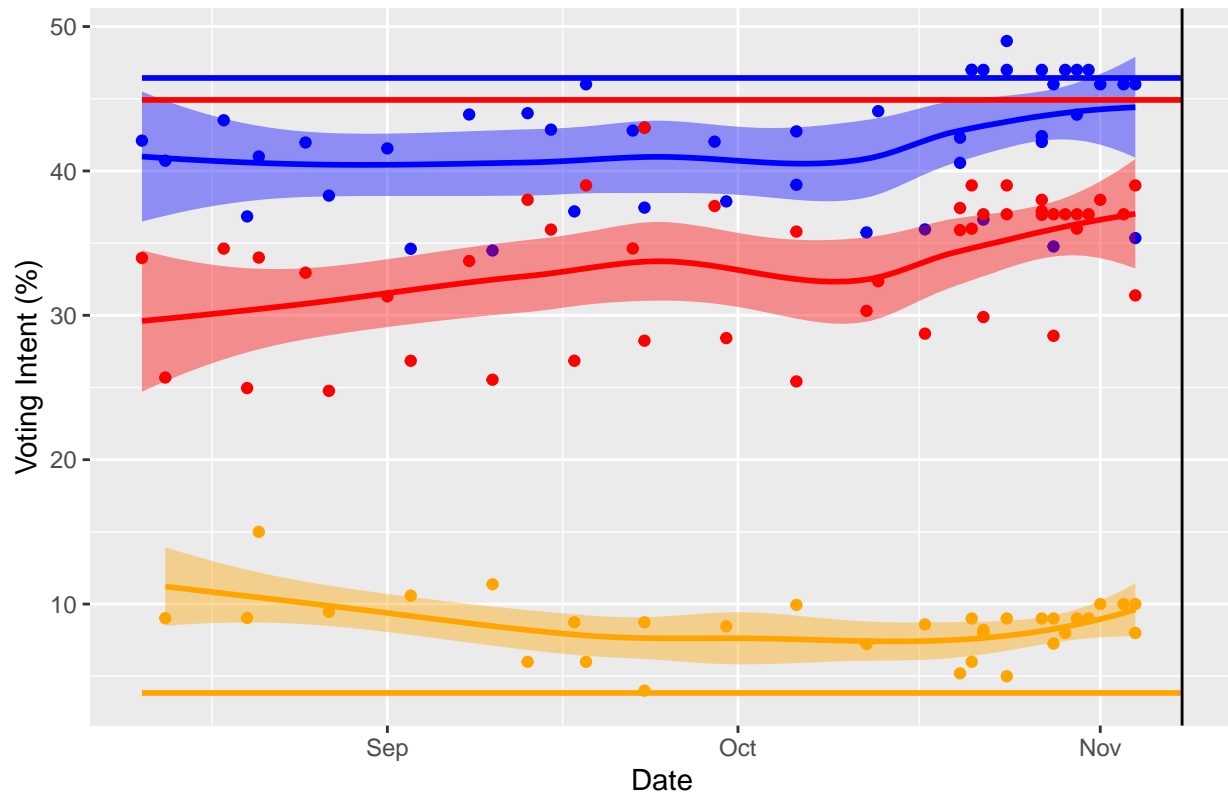
Michigan – Raw Polls



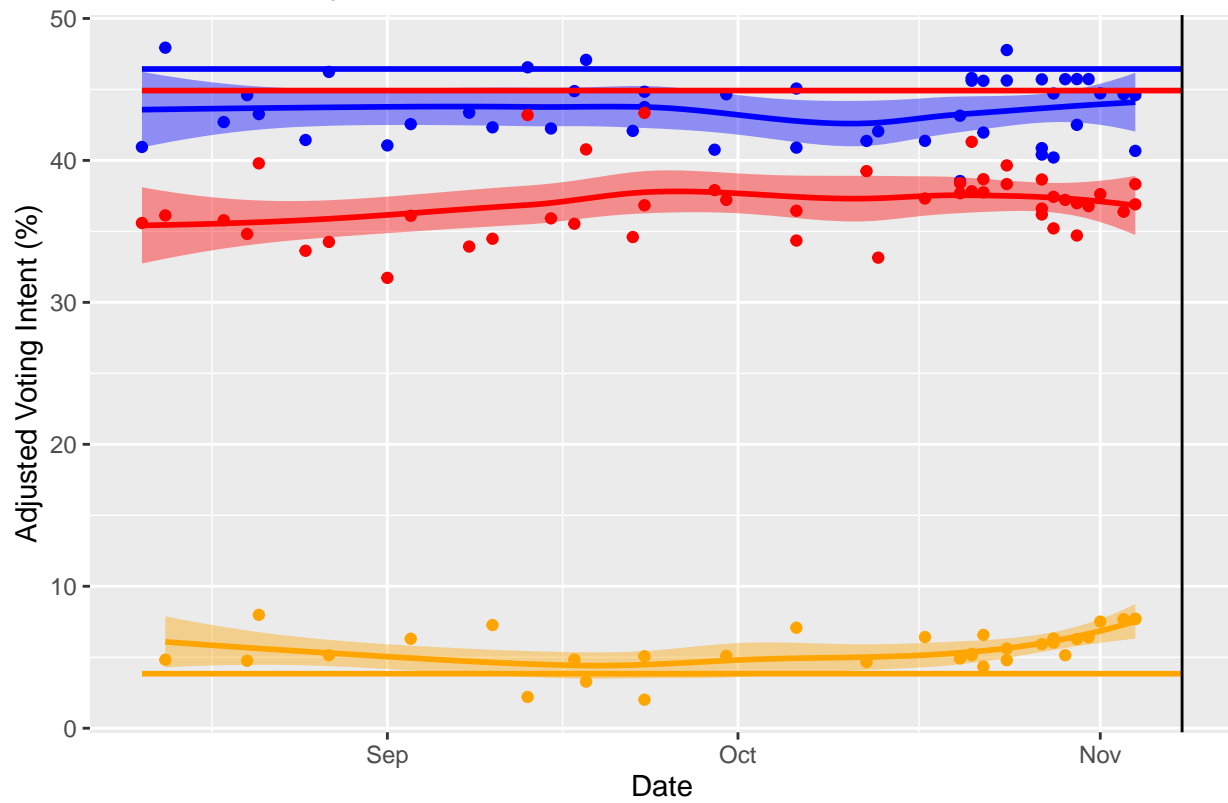
Michigan – Adjusted Polls



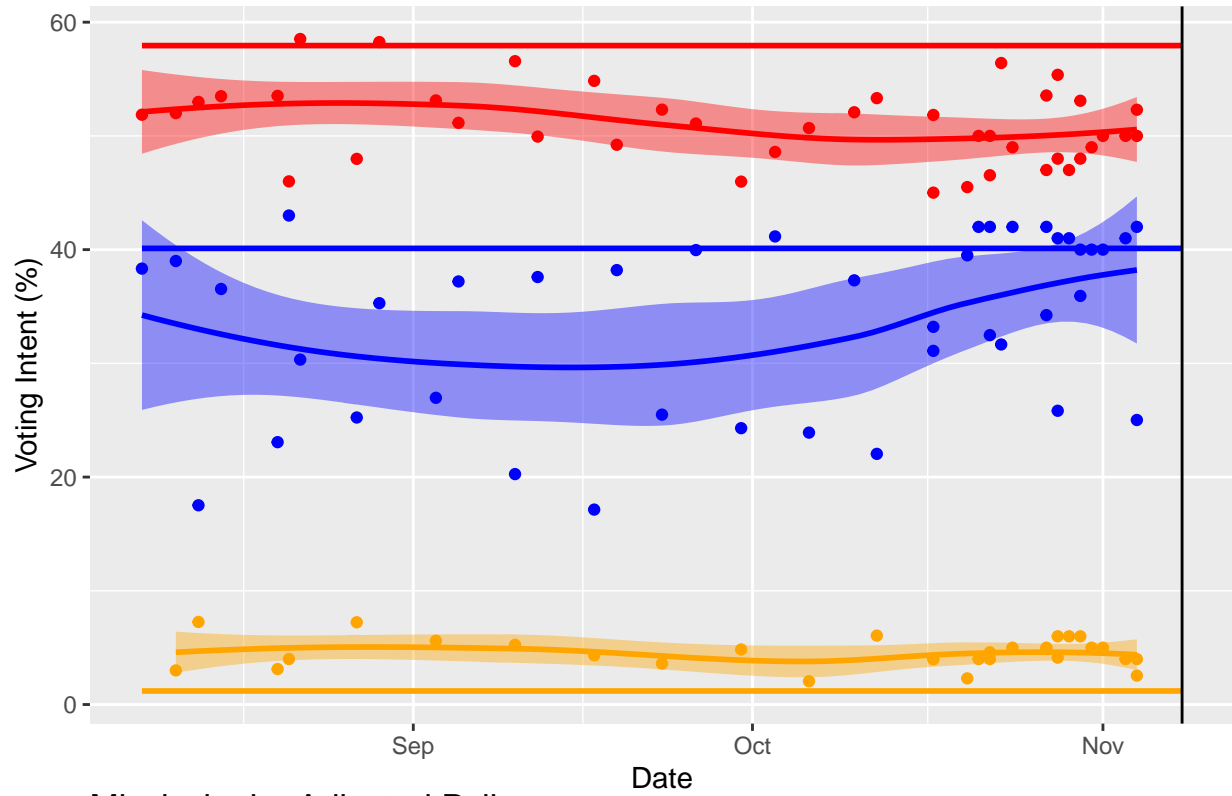
Minnesota – Raw Polls



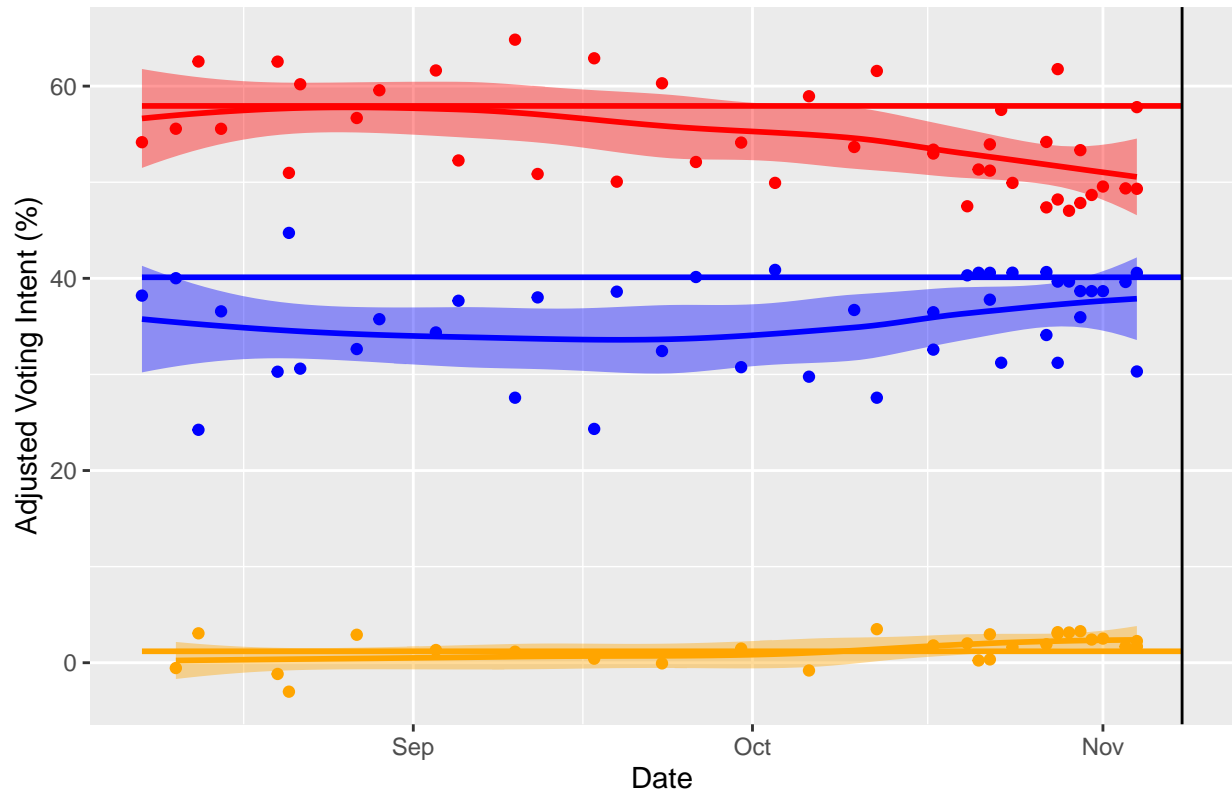
Minnesota – Adjusted Polls



Mississippi – Raw Polls

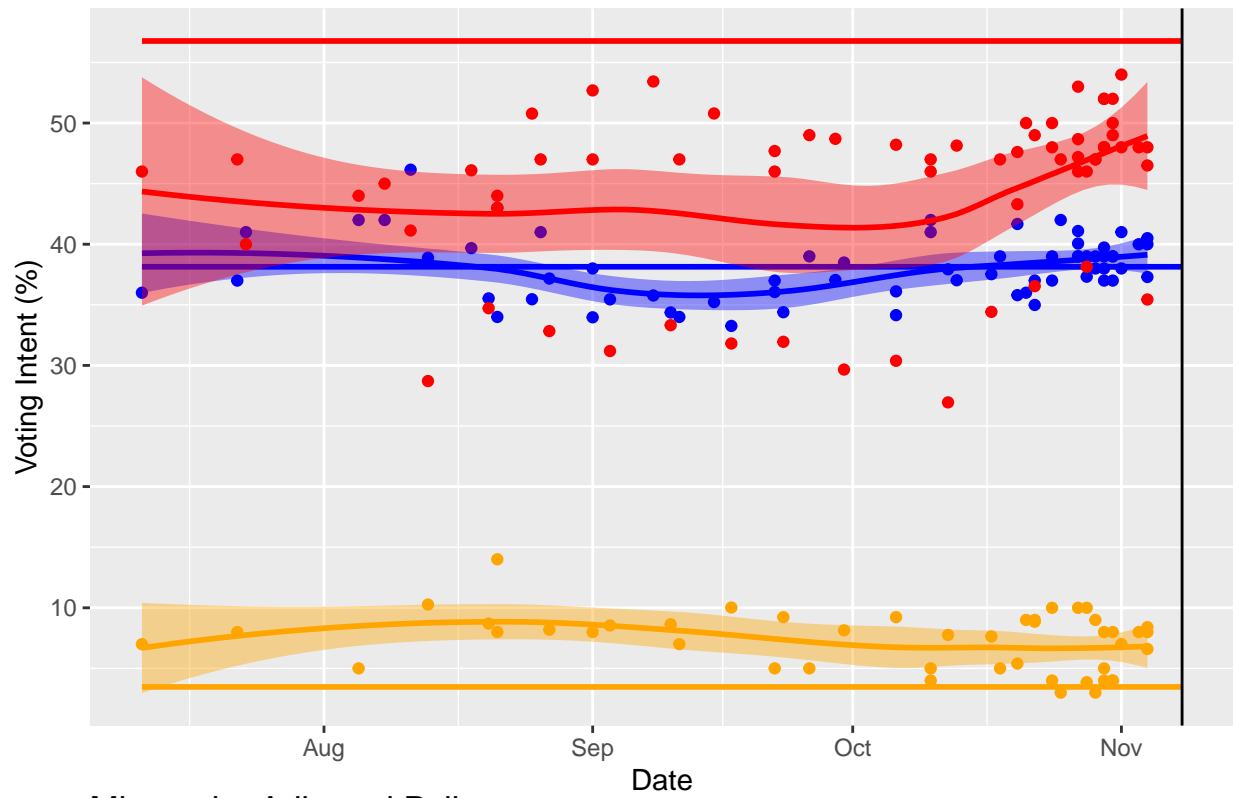


Mississippi – Adjusted Polls

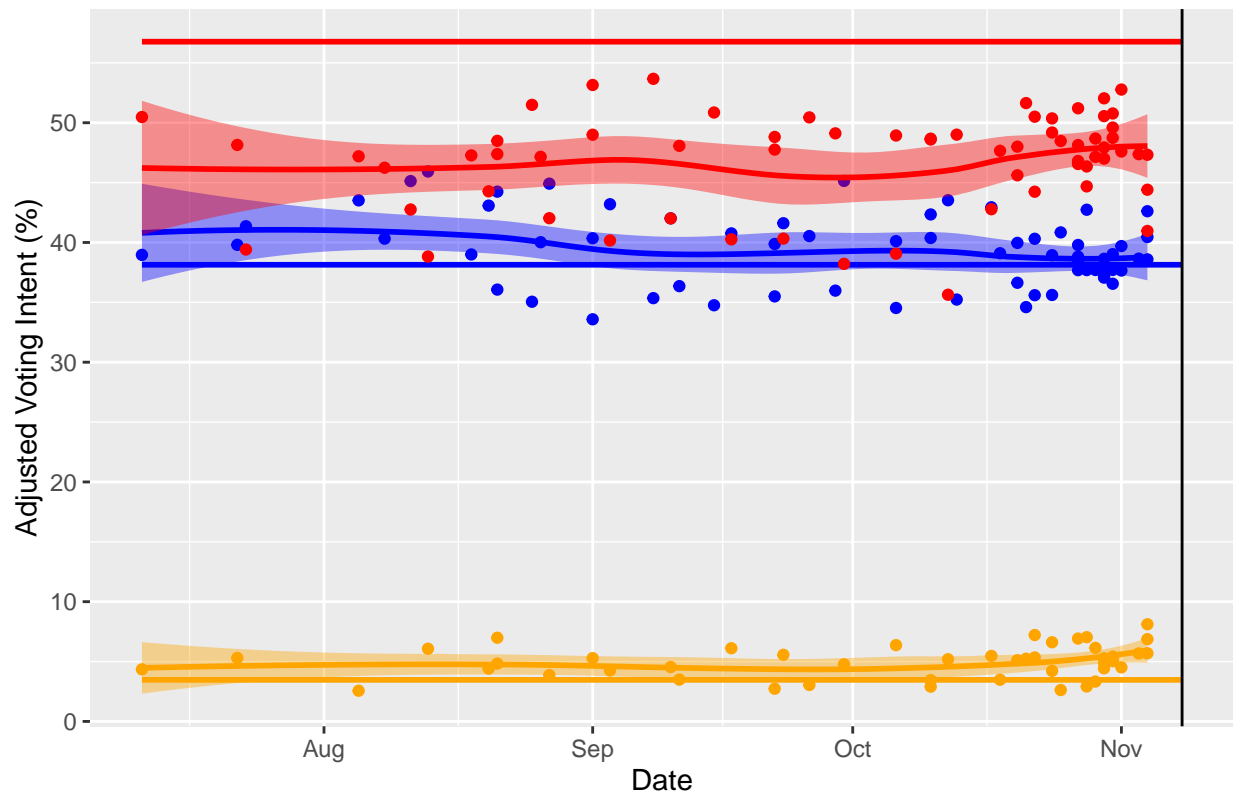




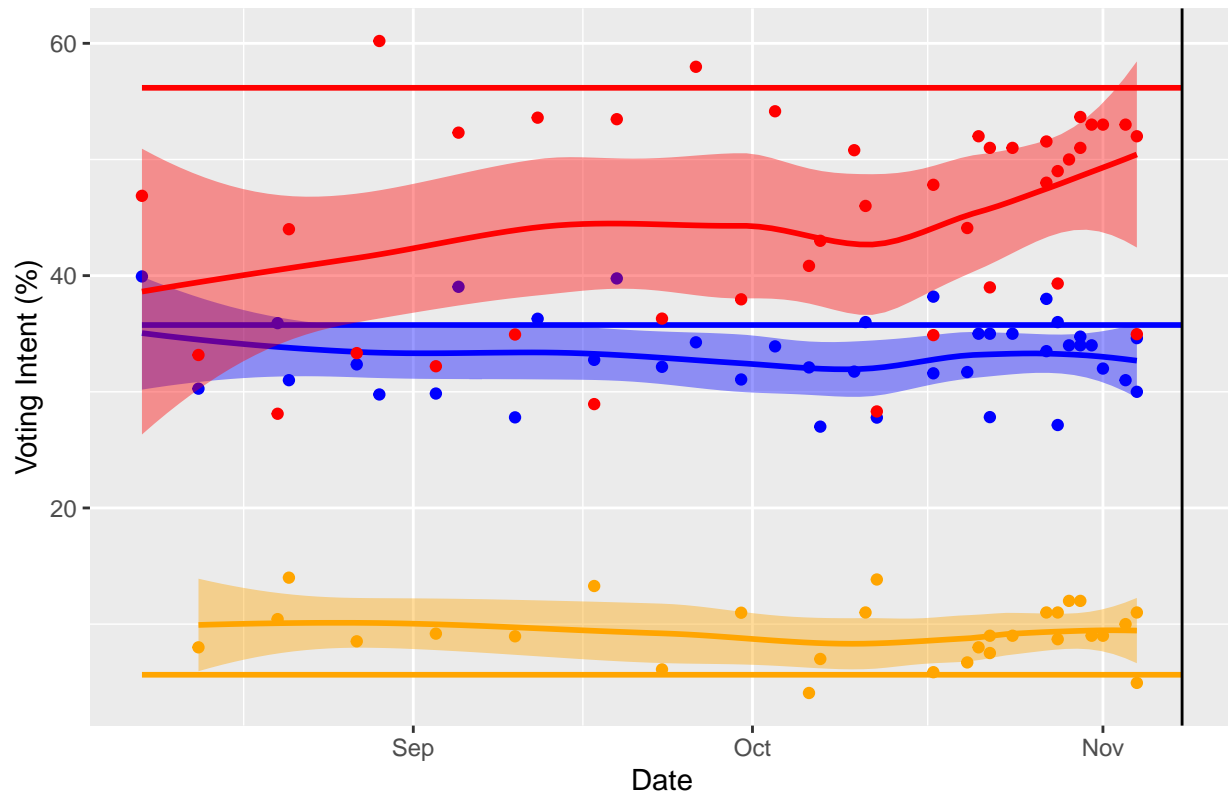
Missouri – Raw Polls



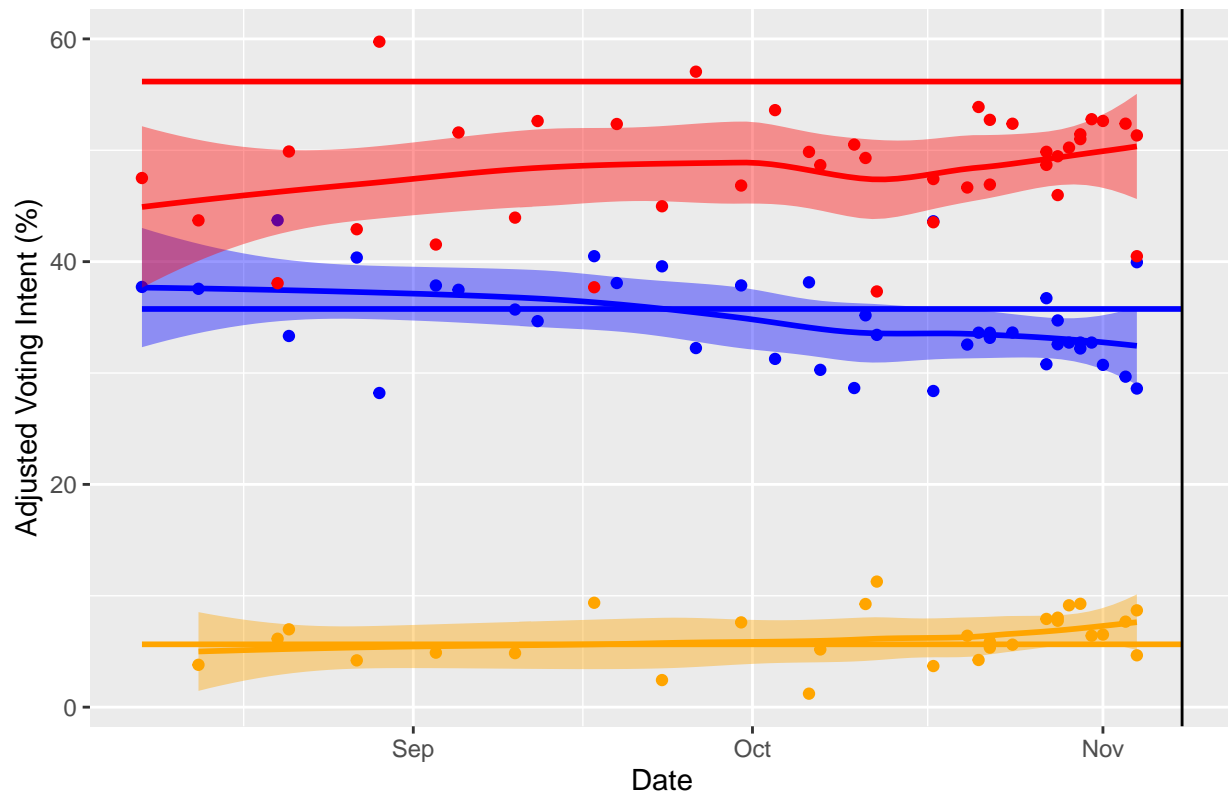
Missouri – Adjusted Polls



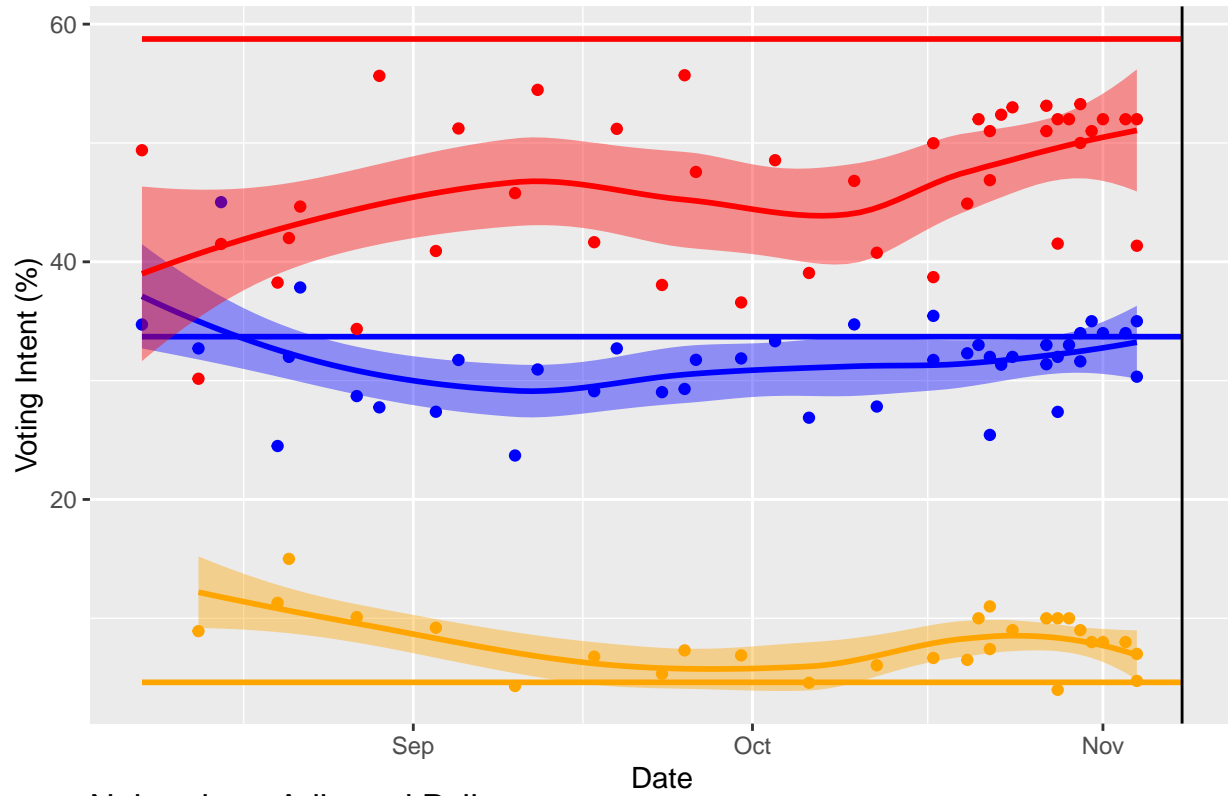
Montana – Raw Polls



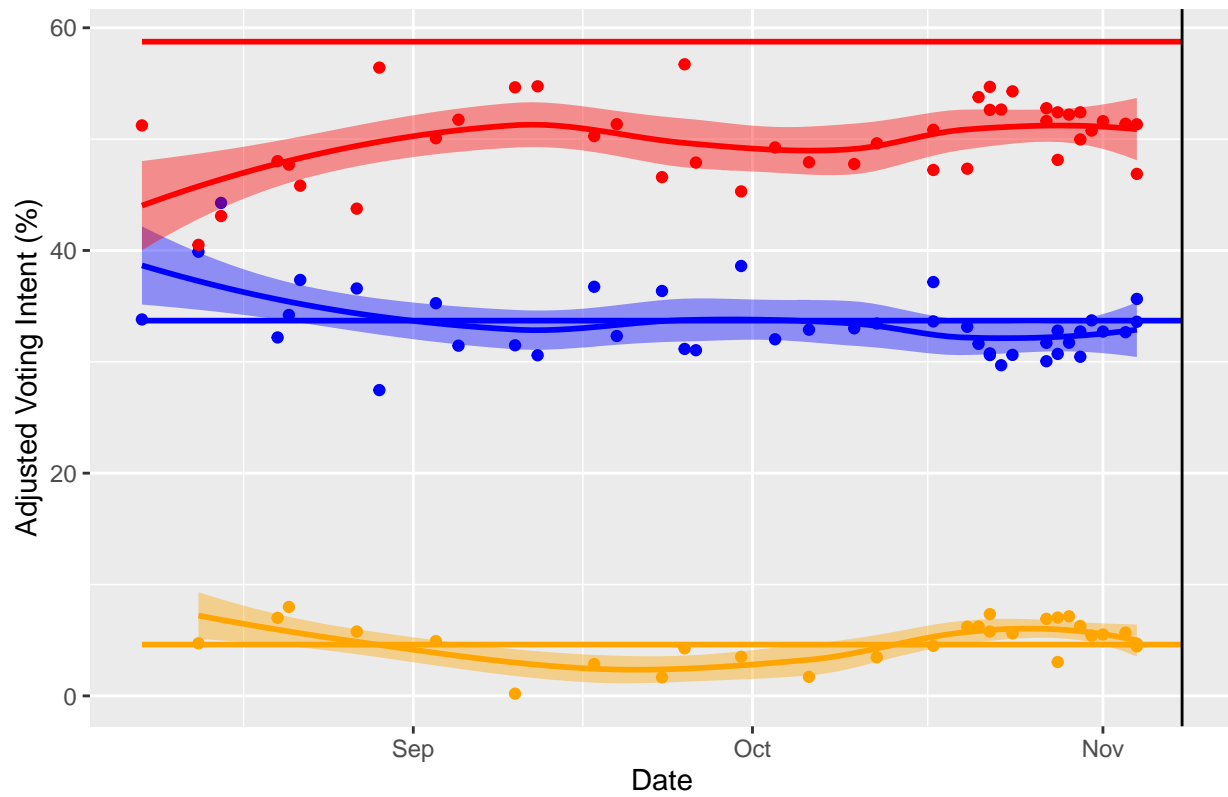
Montana – Adjusted Polls



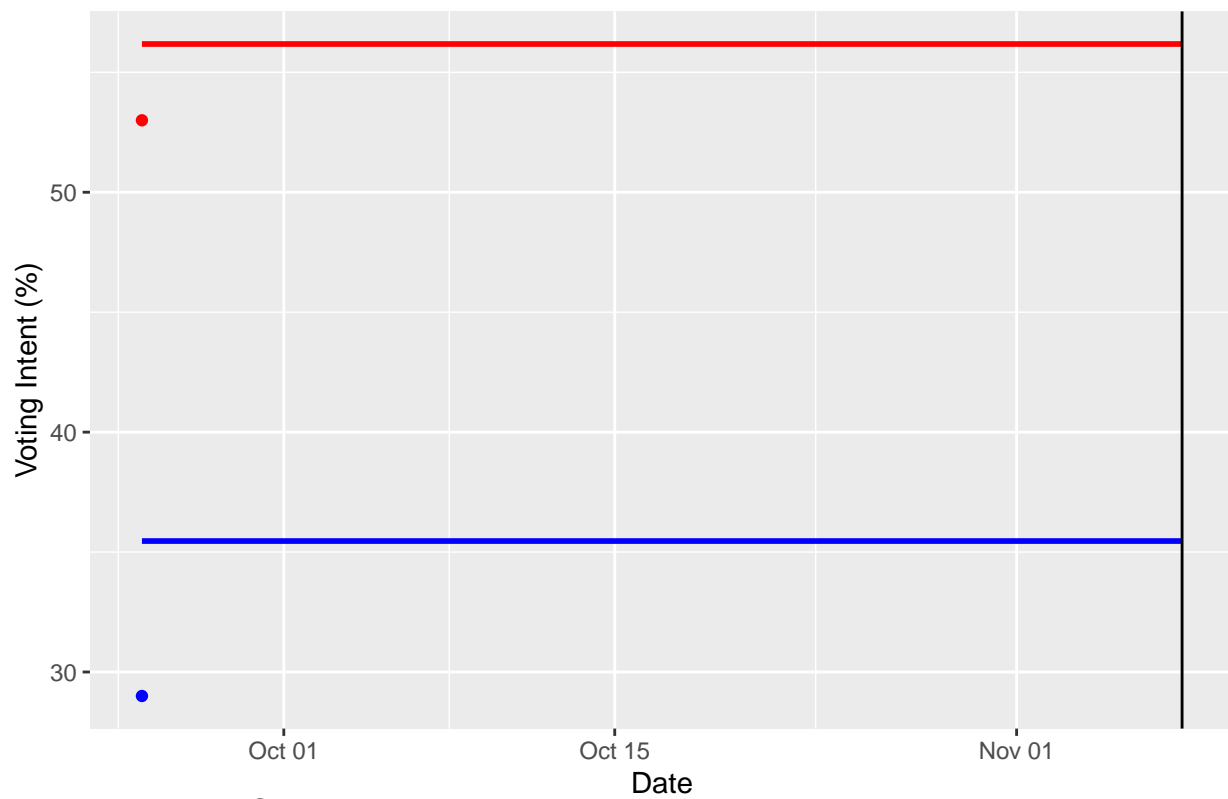
Nebraska – Raw Polls



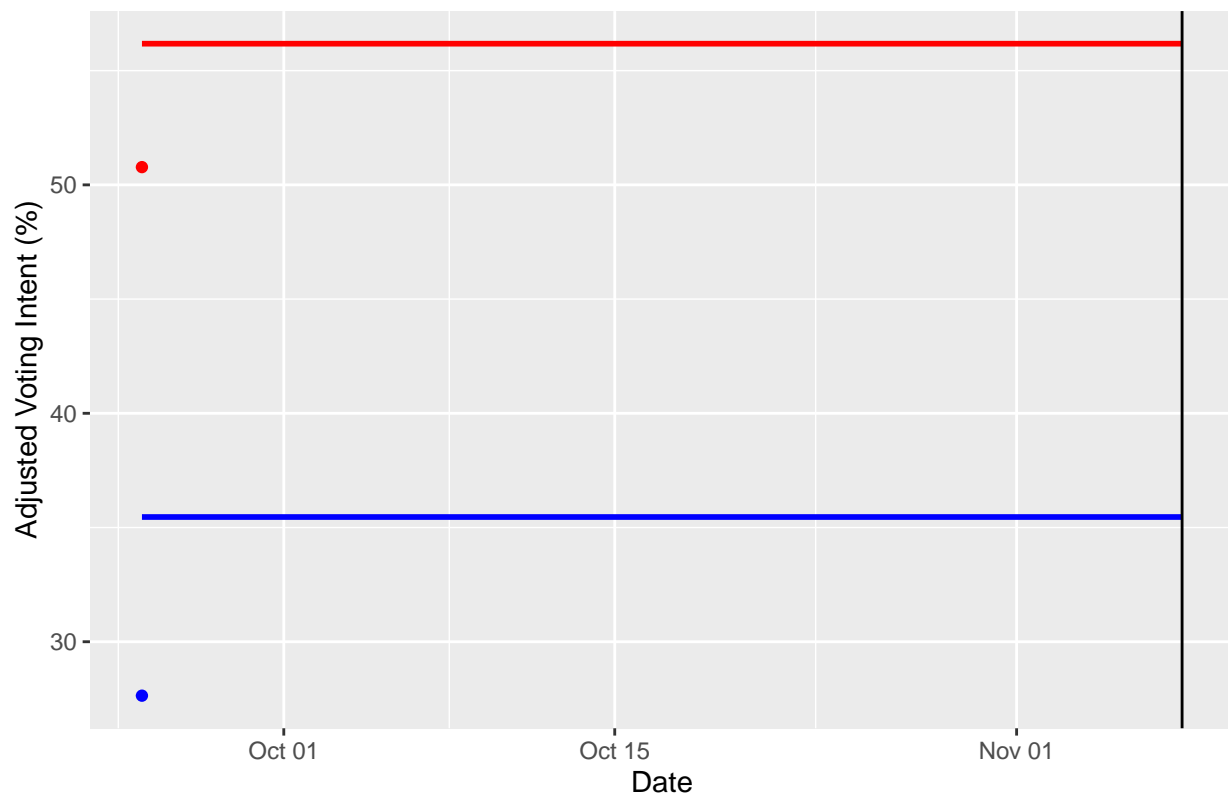
Nebraska – Adjusted Polls



Nebraska CD-1 – Raw Polls



Nebraska CD-1 – Adjusted Polls



```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
```

```

## : span too small.  fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17069

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at 17069

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 0.14

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17097

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : There are other near singularities as well. 0.0196

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger

## Warning: Computation failed in 'stat_smooth()'
## Caused by error in 'predLoess()':
## ! NA/NaN/Inf in foreign function call (arg 5)

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : span too small.  fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17069

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196

```

```

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at 17069

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 0.14

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17097

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger

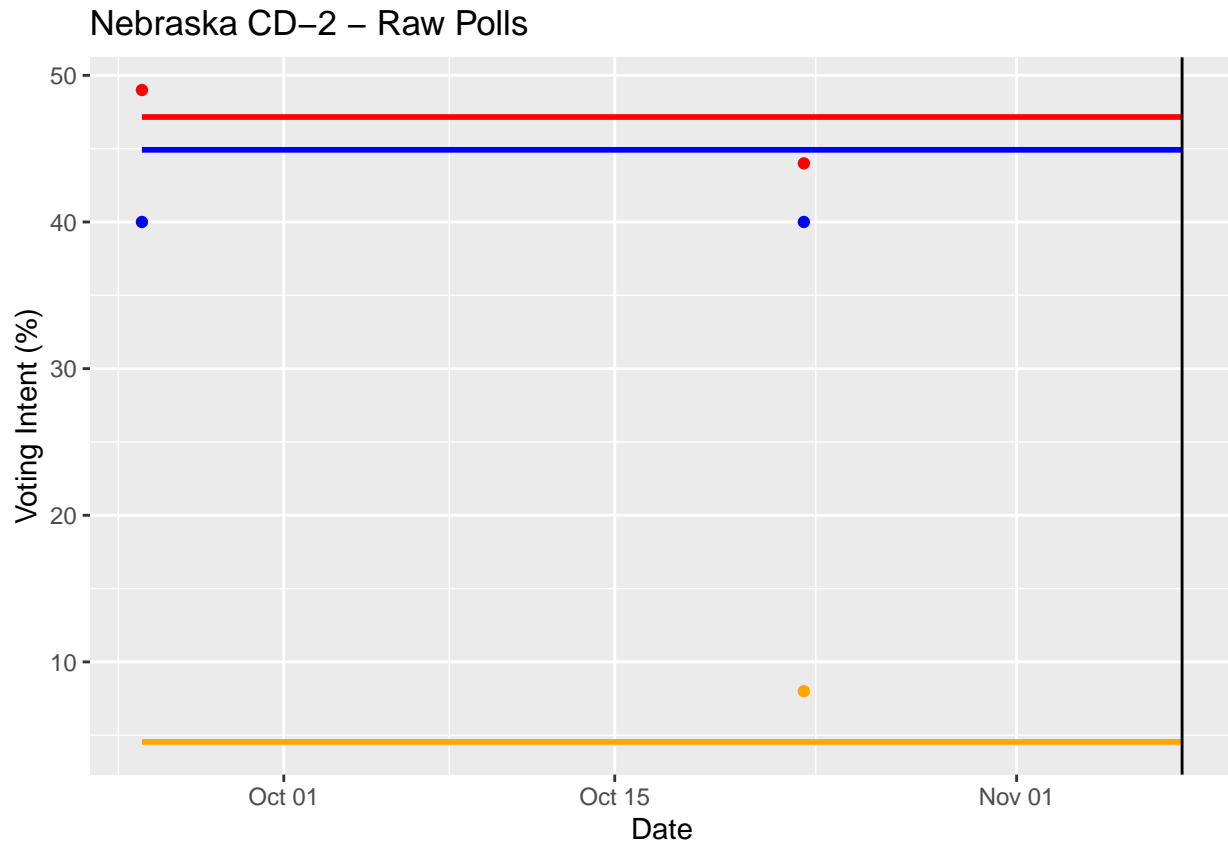
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : There are other near singularities as well. 0.0196

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger

## Warning: Computation failed in 'stat_smooth()'
## Caused by error in 'predLoess()':
## ! NA/NaN/Inf in foreign function call (arg 5)

```



```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : span too small. fewer data values than degrees of freedom.
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17069
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at 17069
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 0.14
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 1
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17097
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196
```

```

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : There are other near singularities as well. 0.0196

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : zero-width neighborhood. make span bigger

## Warning: Computation failed in 'stat_smooth()'
## Caused by error in 'predLoess()':
## ! NA/NaN/Inf in foreign function call (arg 5)

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17069

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at 17069

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 0.14

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : at 17097

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : radius 0.0196

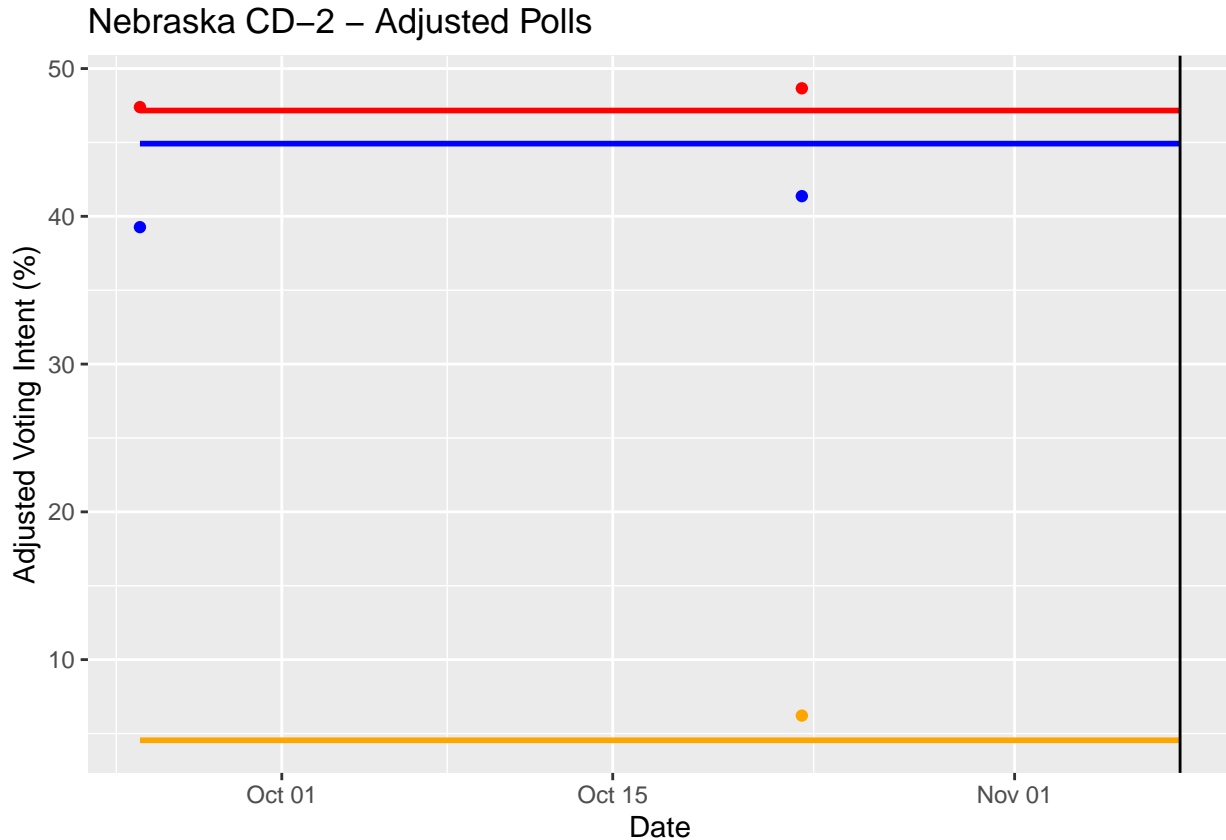
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : There are other near singularities as well. 0.0196

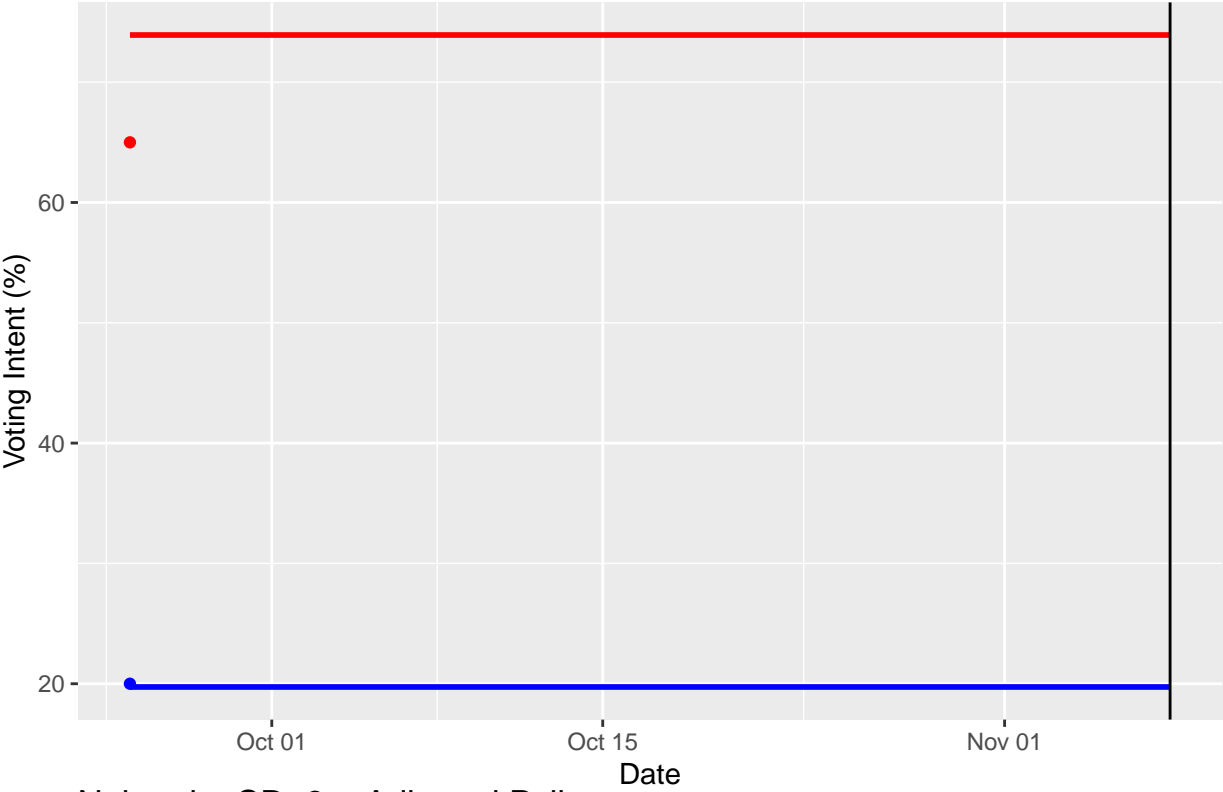
```



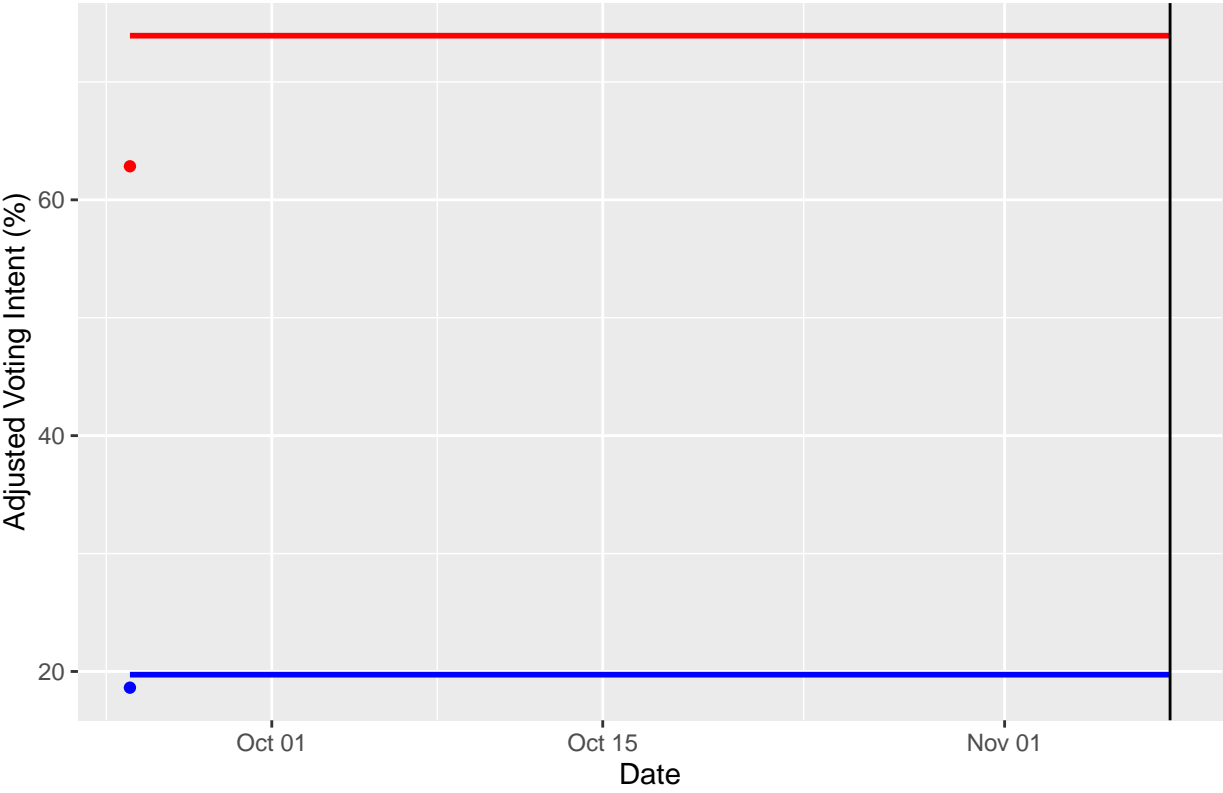
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : zero-width neighborhood. make span bigger  
  
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : zero-width neighborhood. make span bigger  
  
## Warning: Computation failed in 'stat_smooth()'  
## Caused by error in 'predLoess()':  
## ! NA/NaN/Inf in foreign function call (arg 5)
```



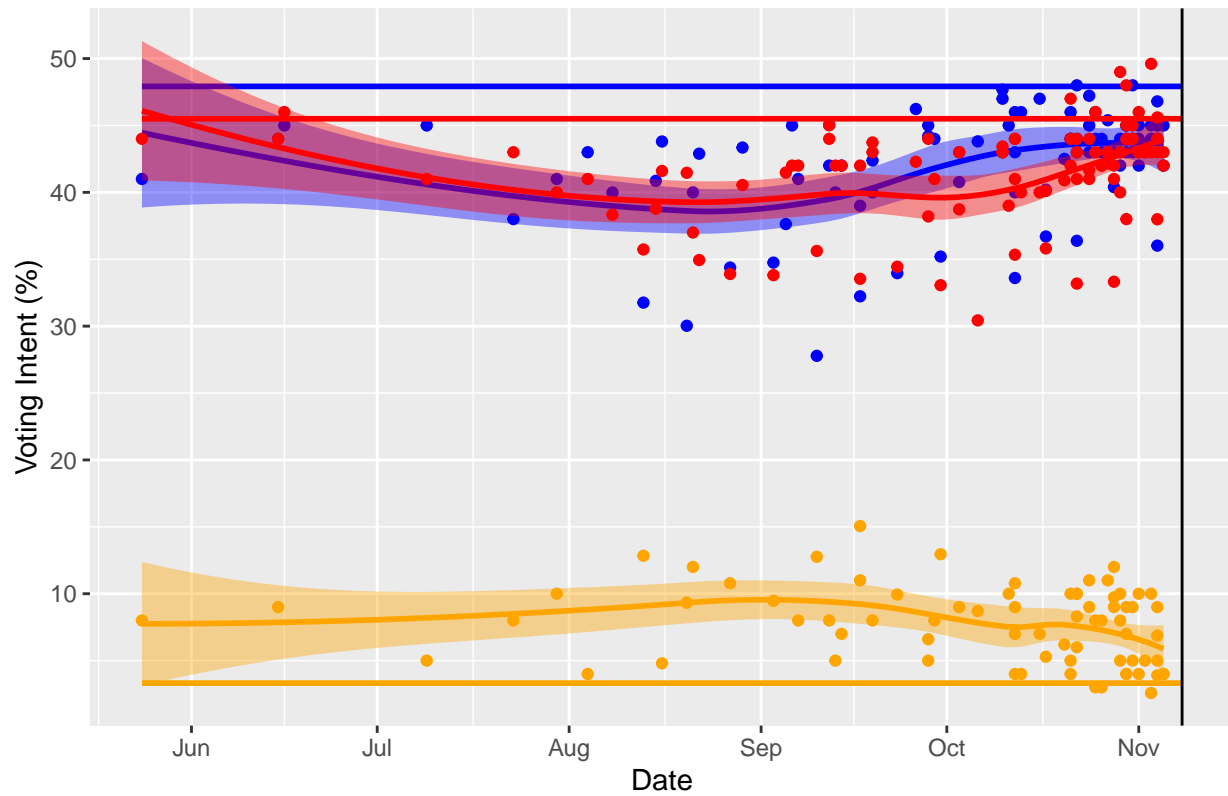
Nebraska CD-3 – Raw Polls



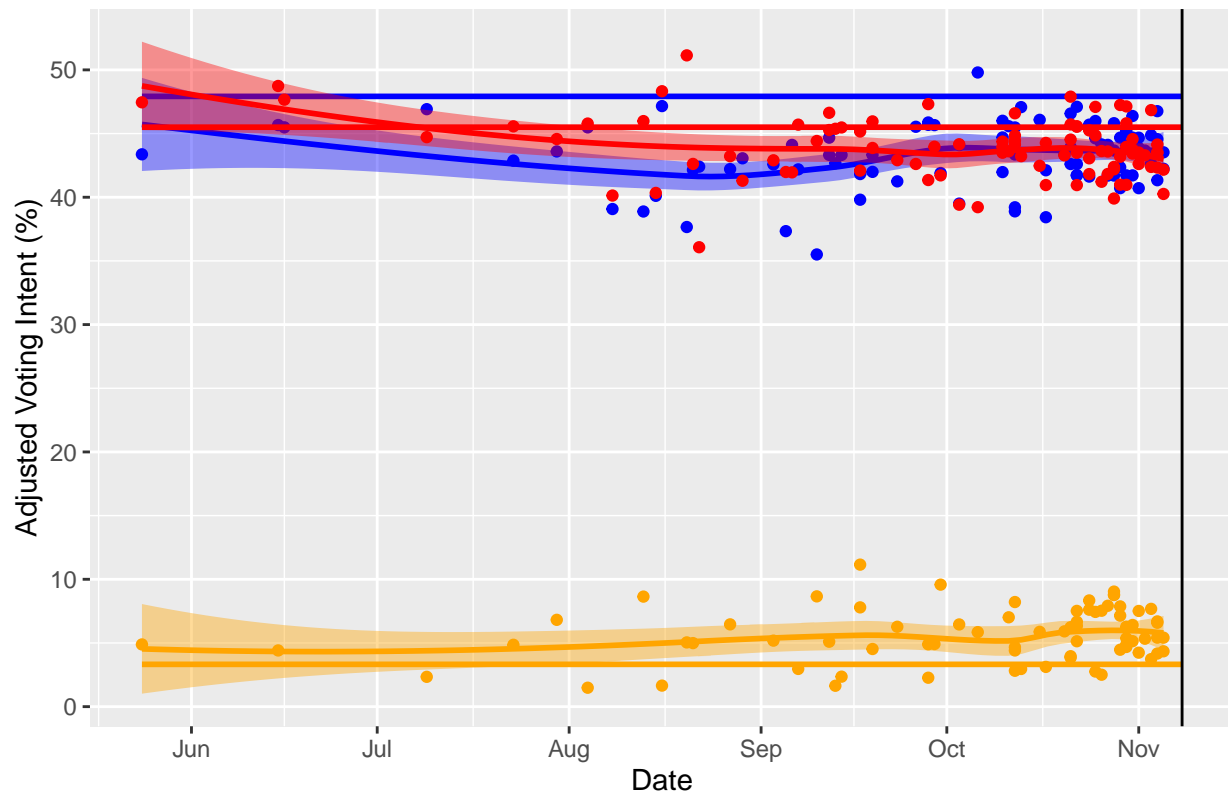
Nebraska CD-3 – Adjusted Polls



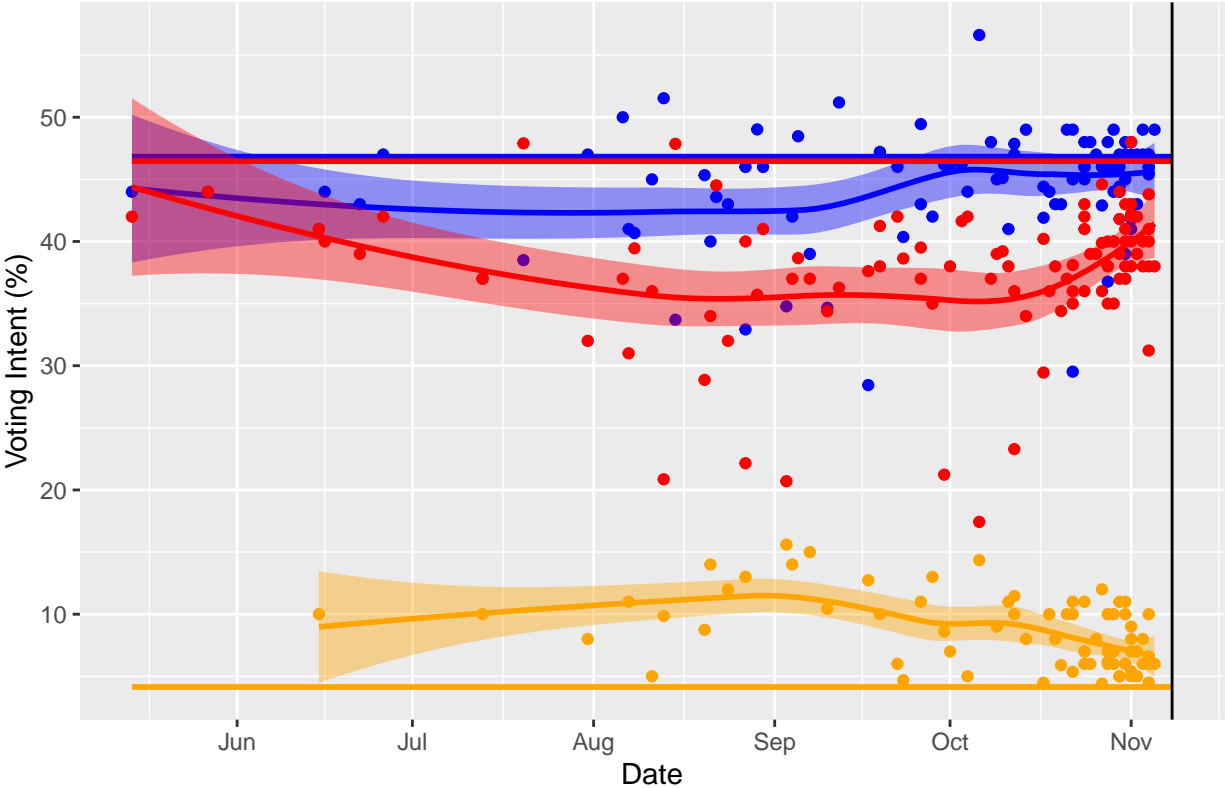
Nevada – Raw Polls



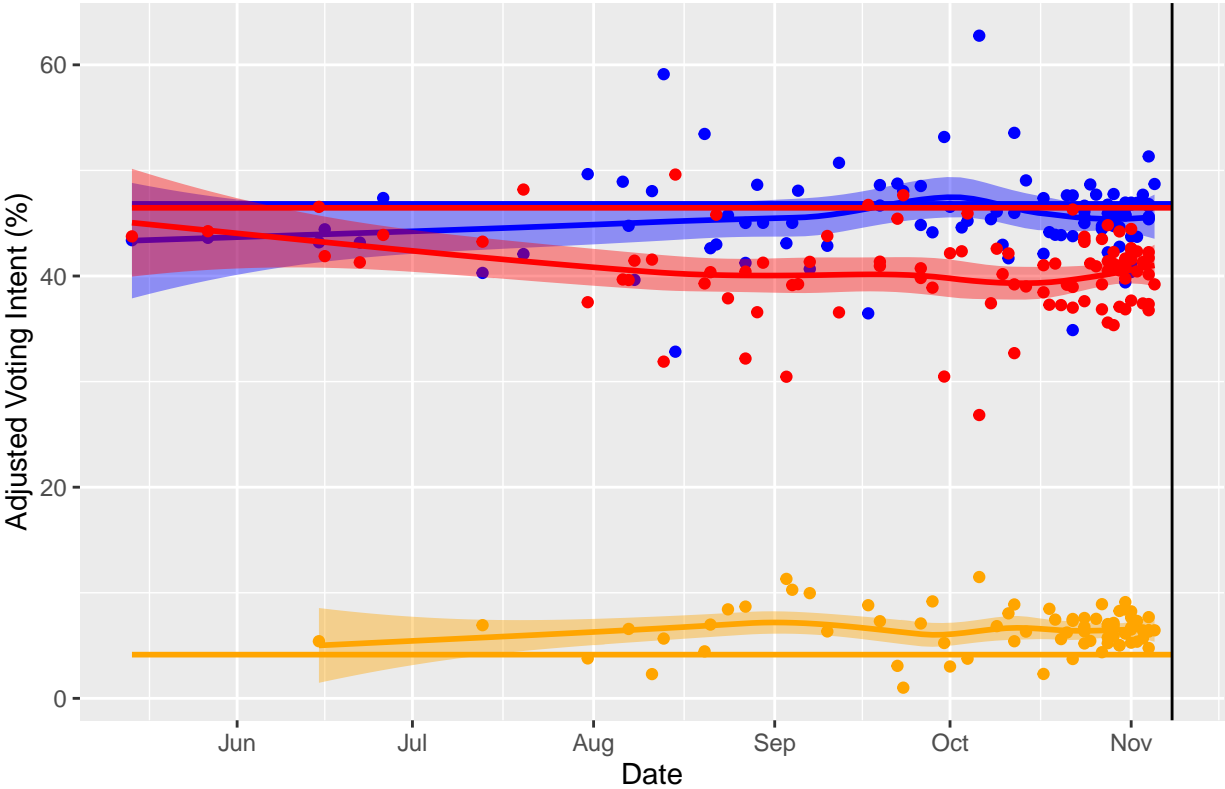
Nevada – Adjusted Polls



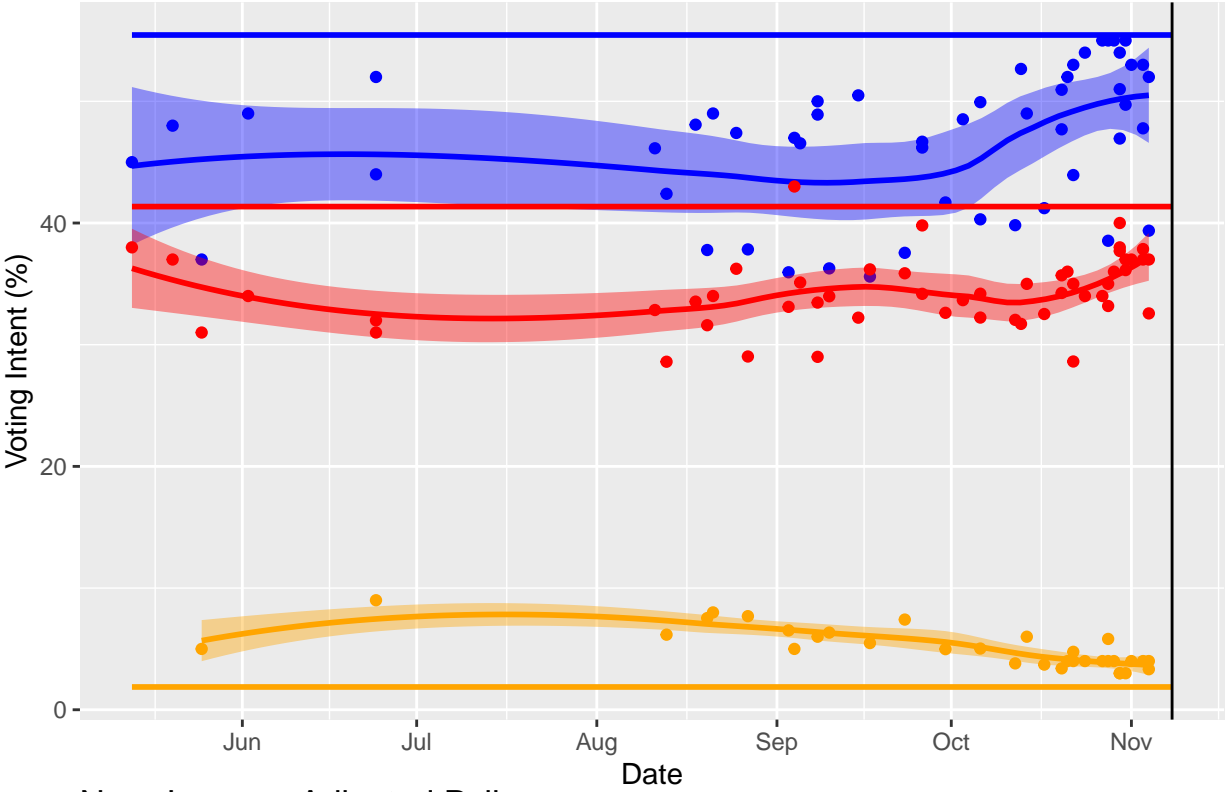
New Hampshire – Raw Polls



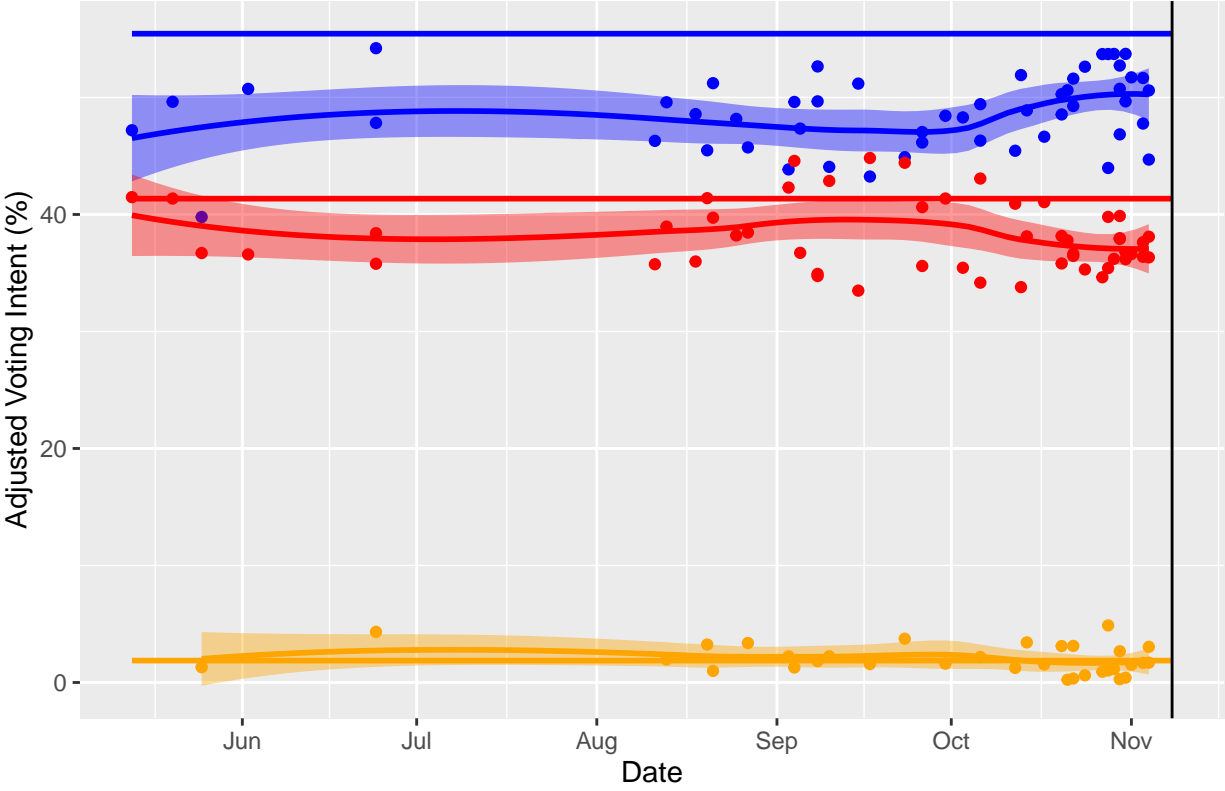
New Hampshire – Adjusted Polls



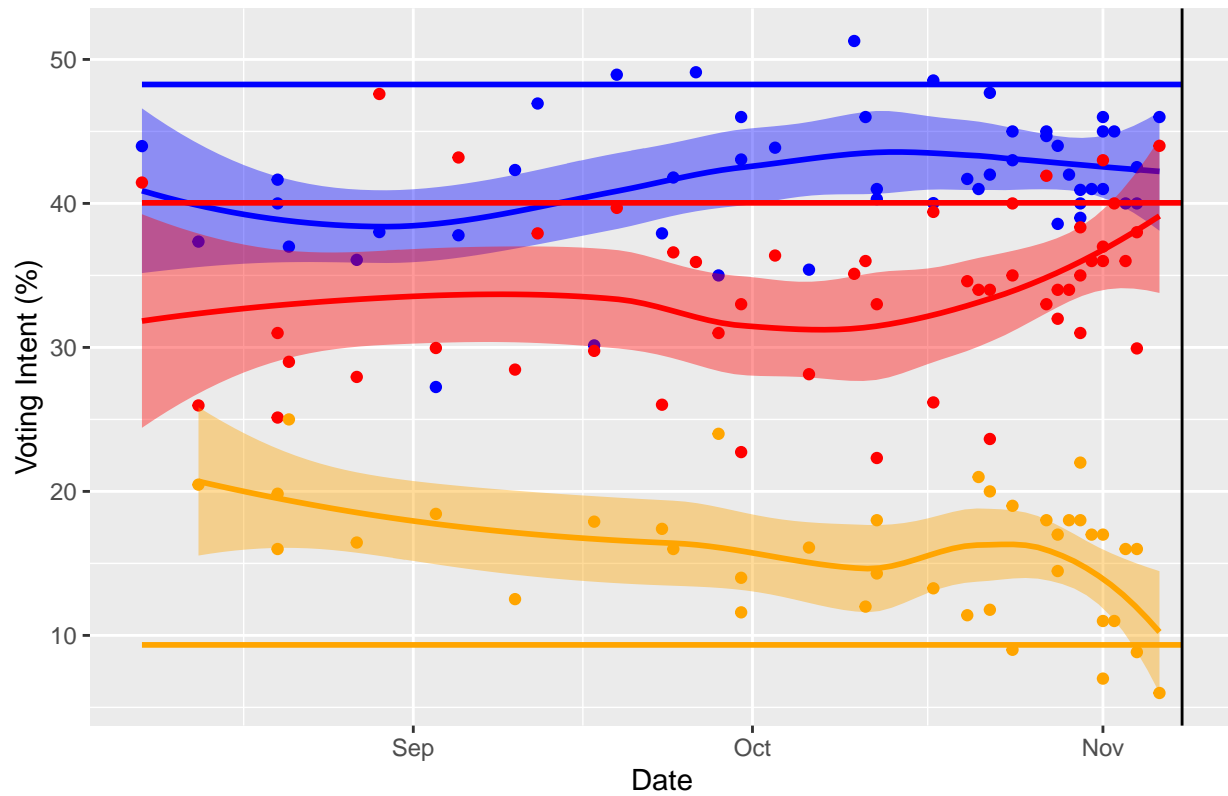
New Jersey – Raw Polls



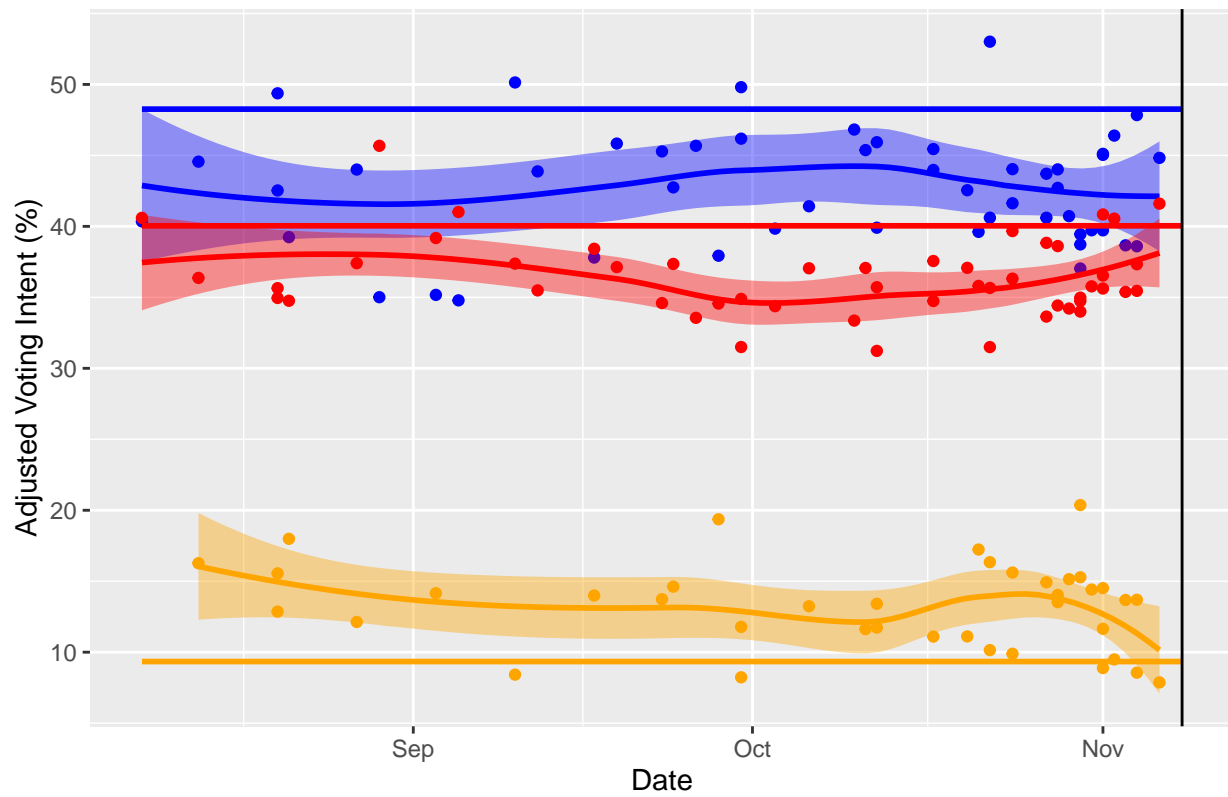
New Jersey – Adjusted Polls



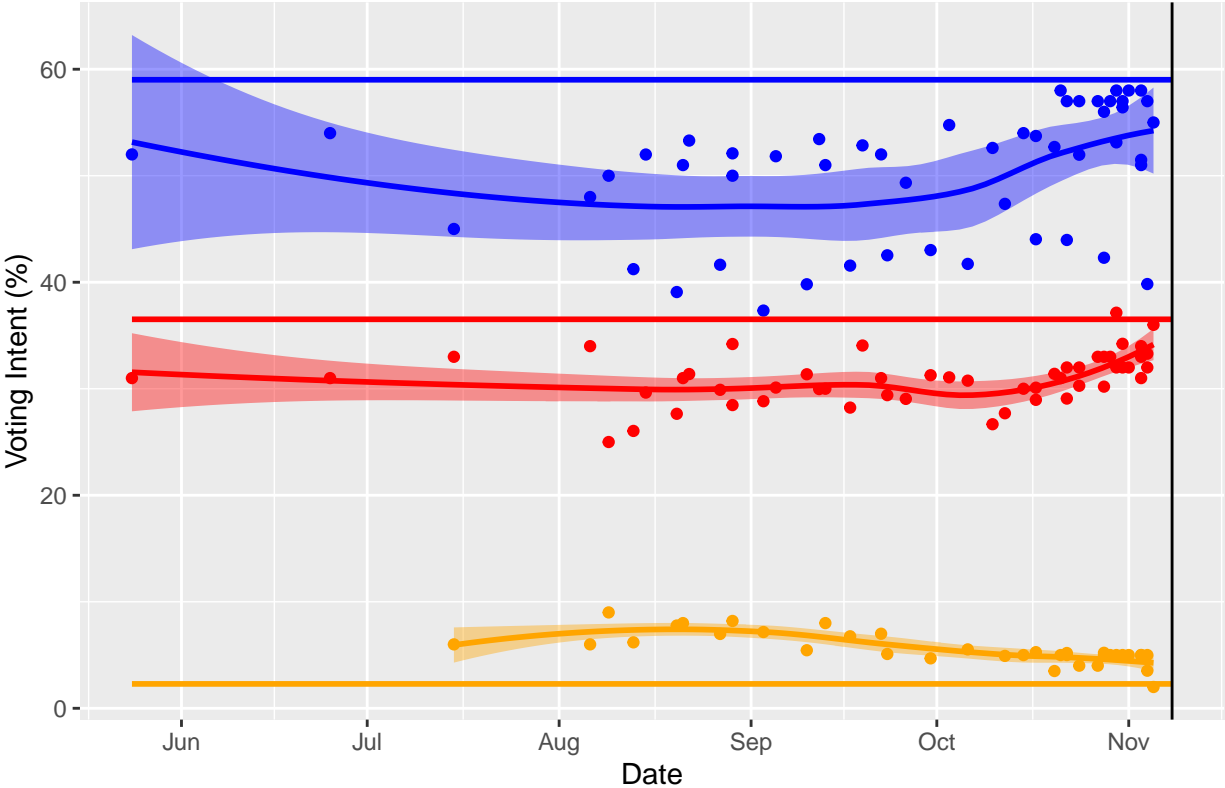
New Mexico – Raw Polls



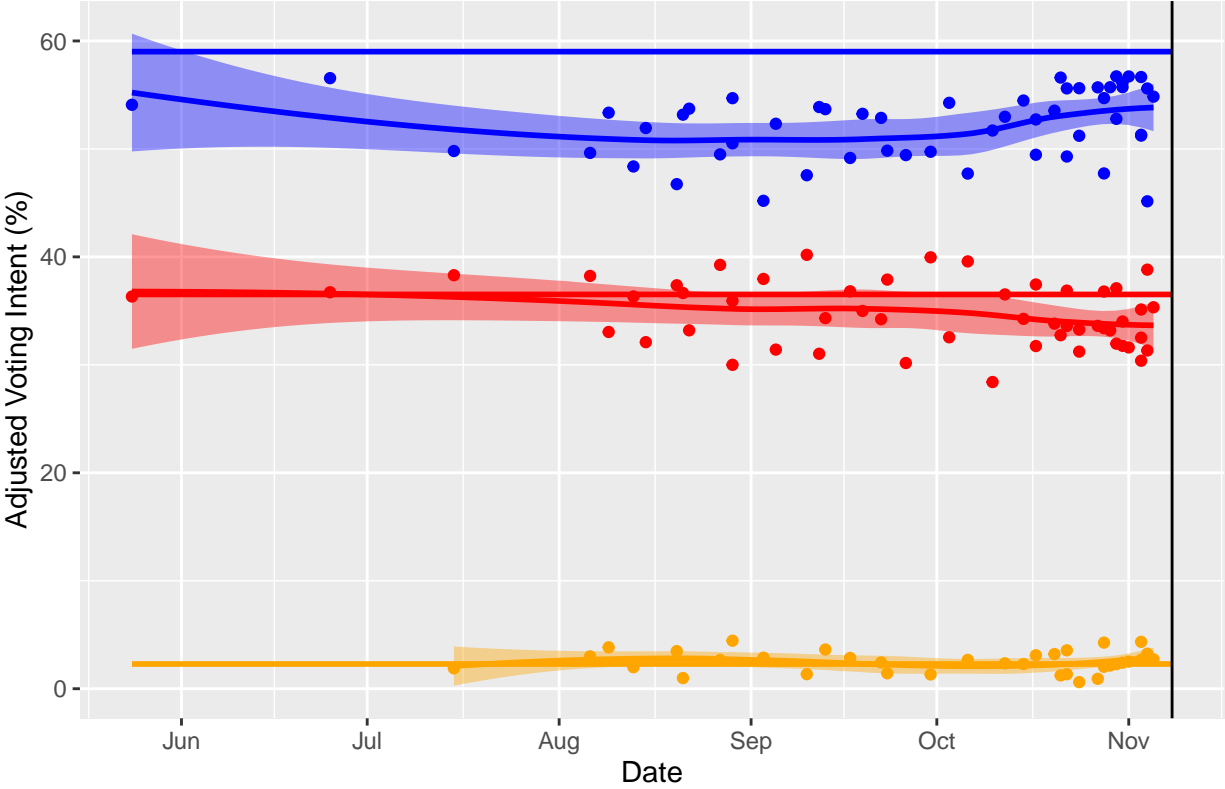
New Mexico – Adjusted Polls



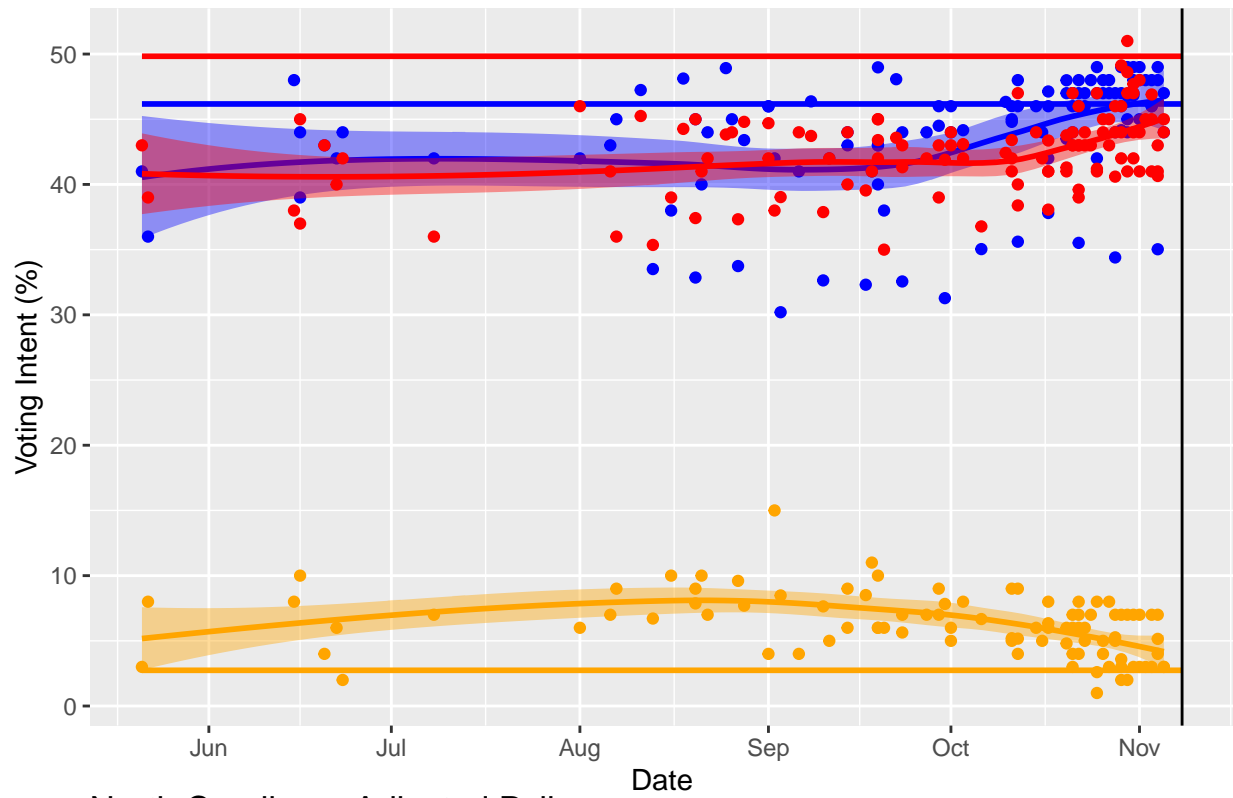
New York – Raw Polls



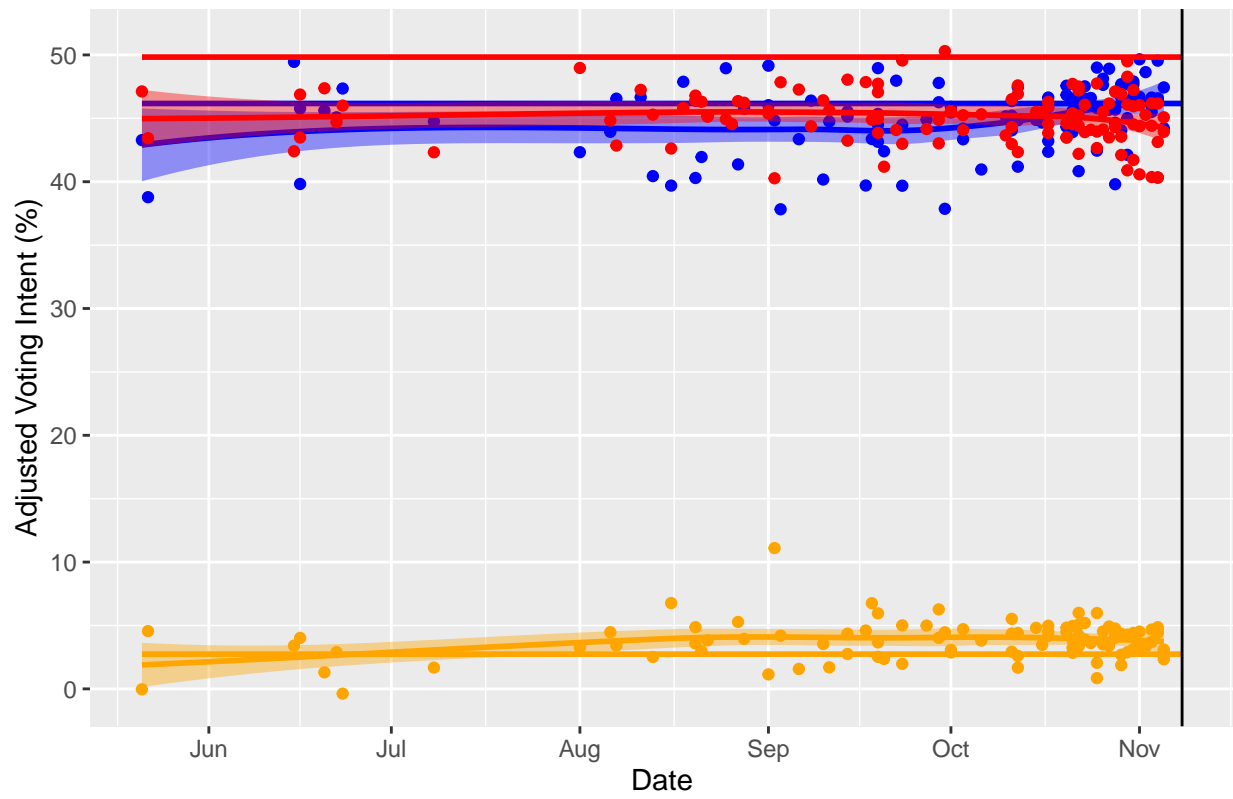
New York – Adjusted Polls



North Carolina – Raw Polls

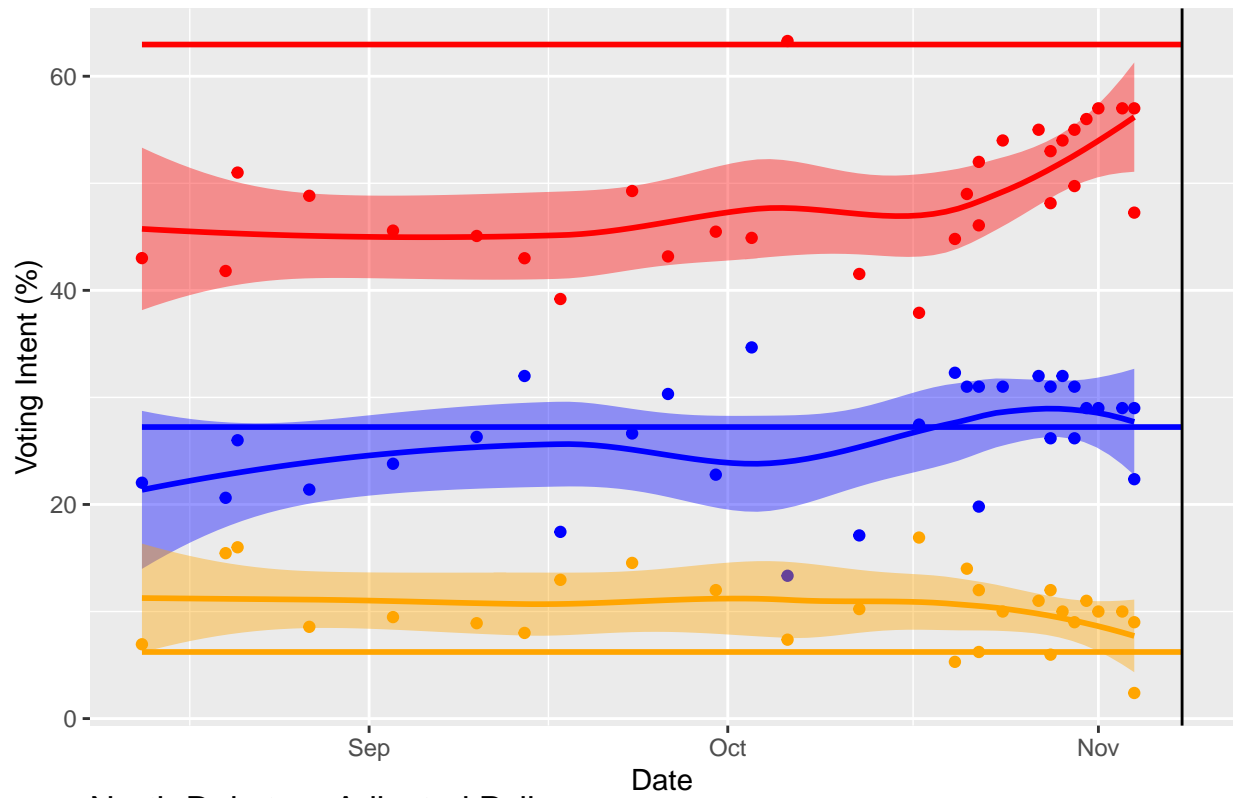


North Carolina – Adjusted Polls

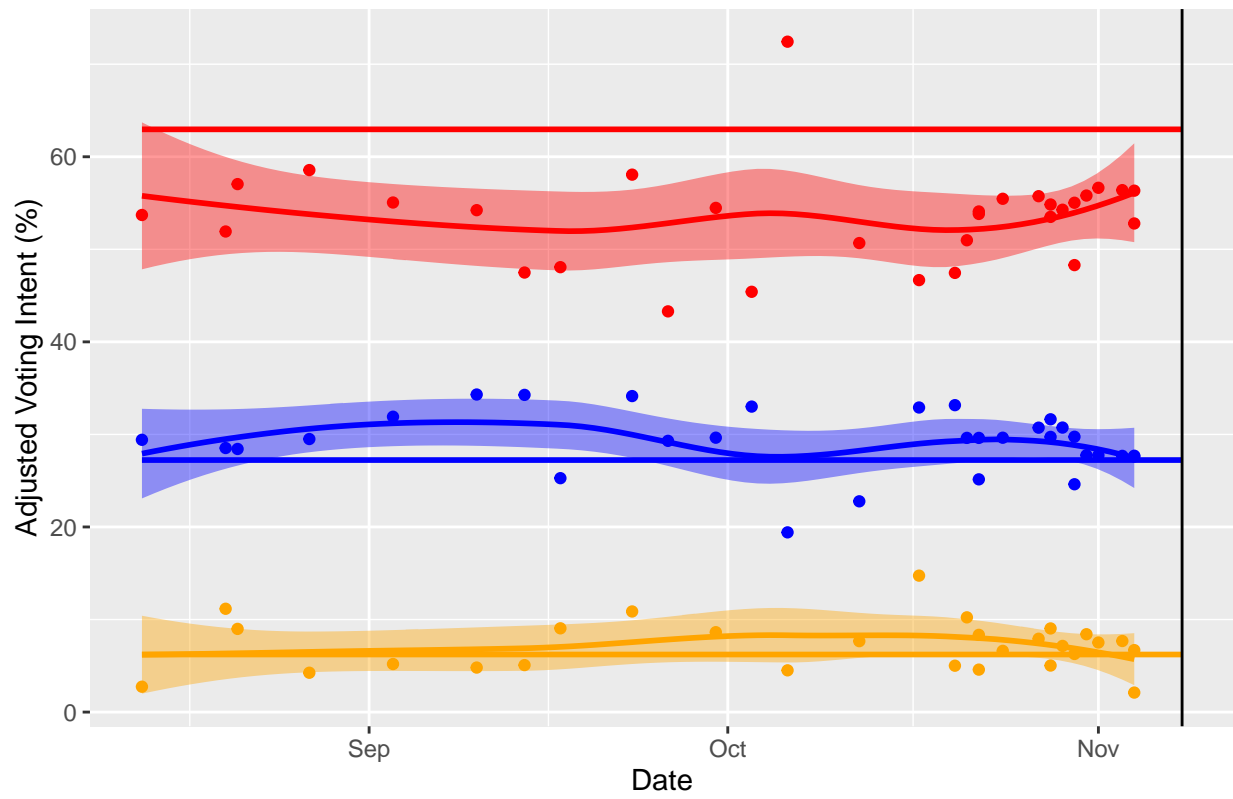




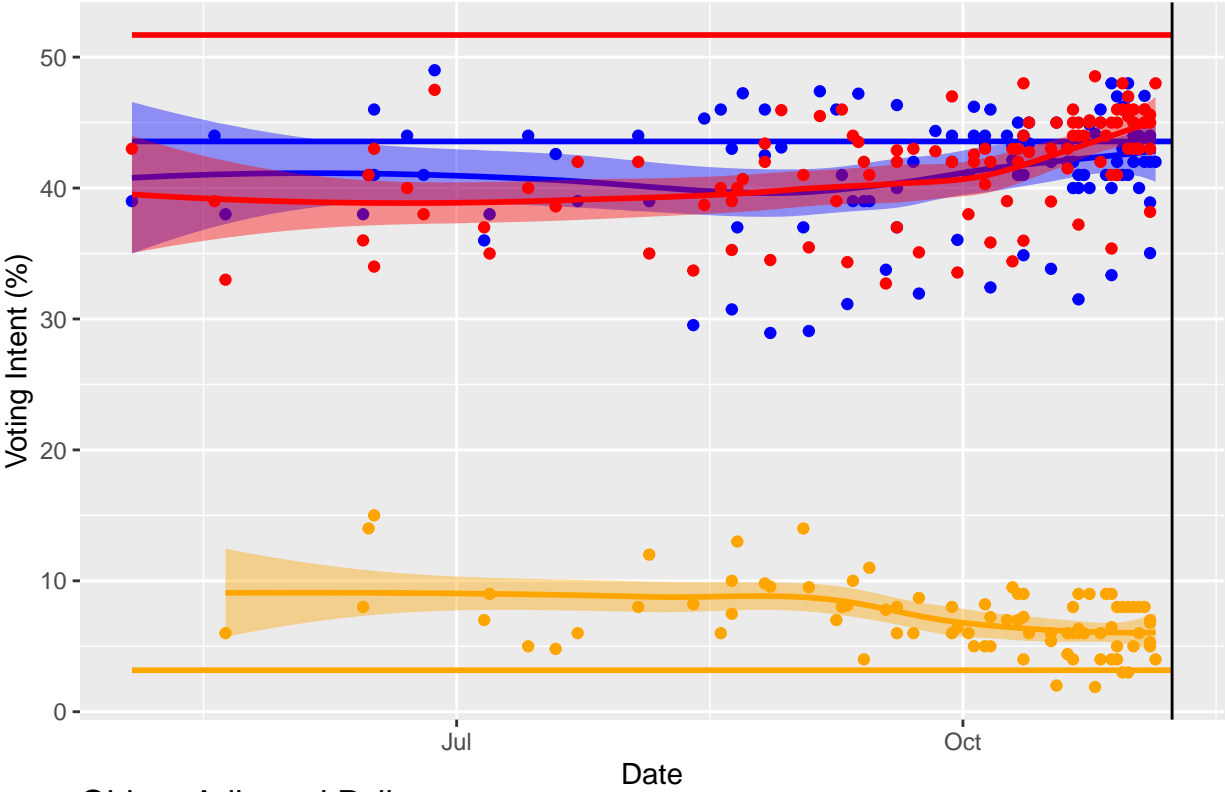
North Dakota – Raw Polls



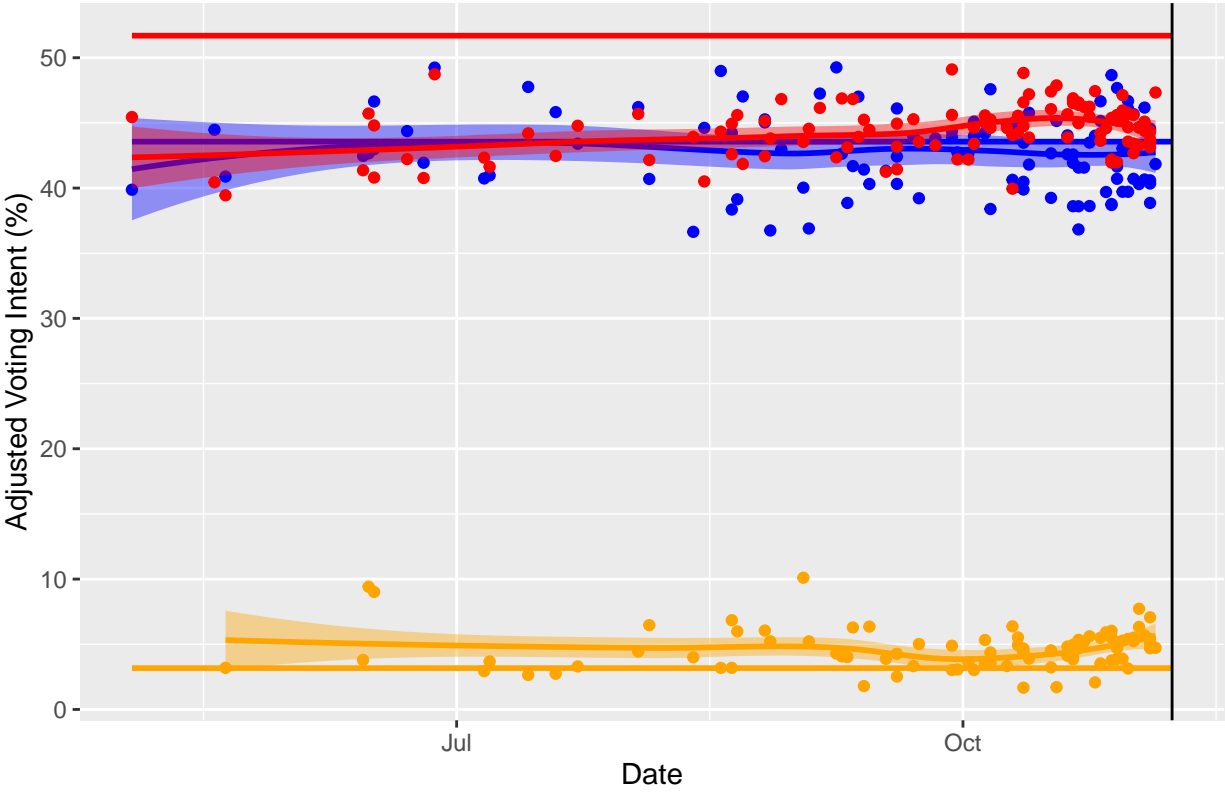
North Dakota – Adjusted Polls



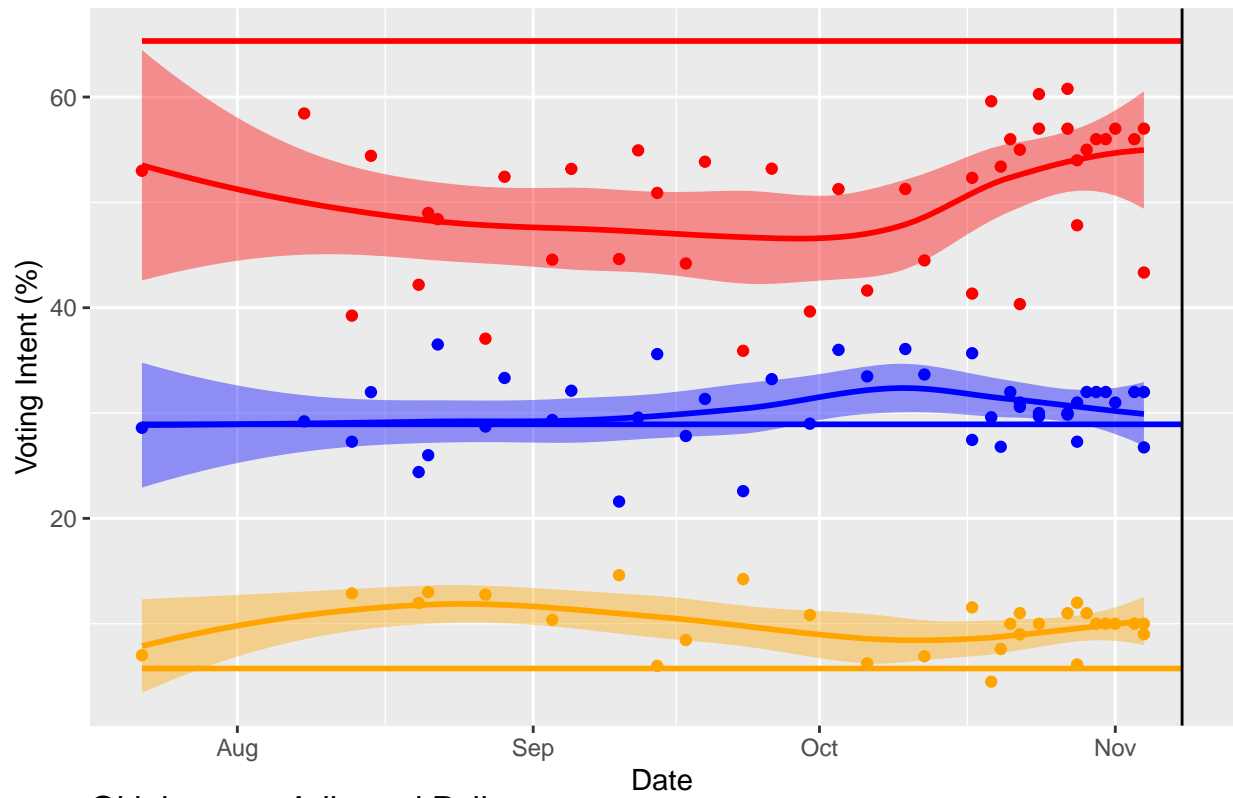
Ohio – Raw Polls



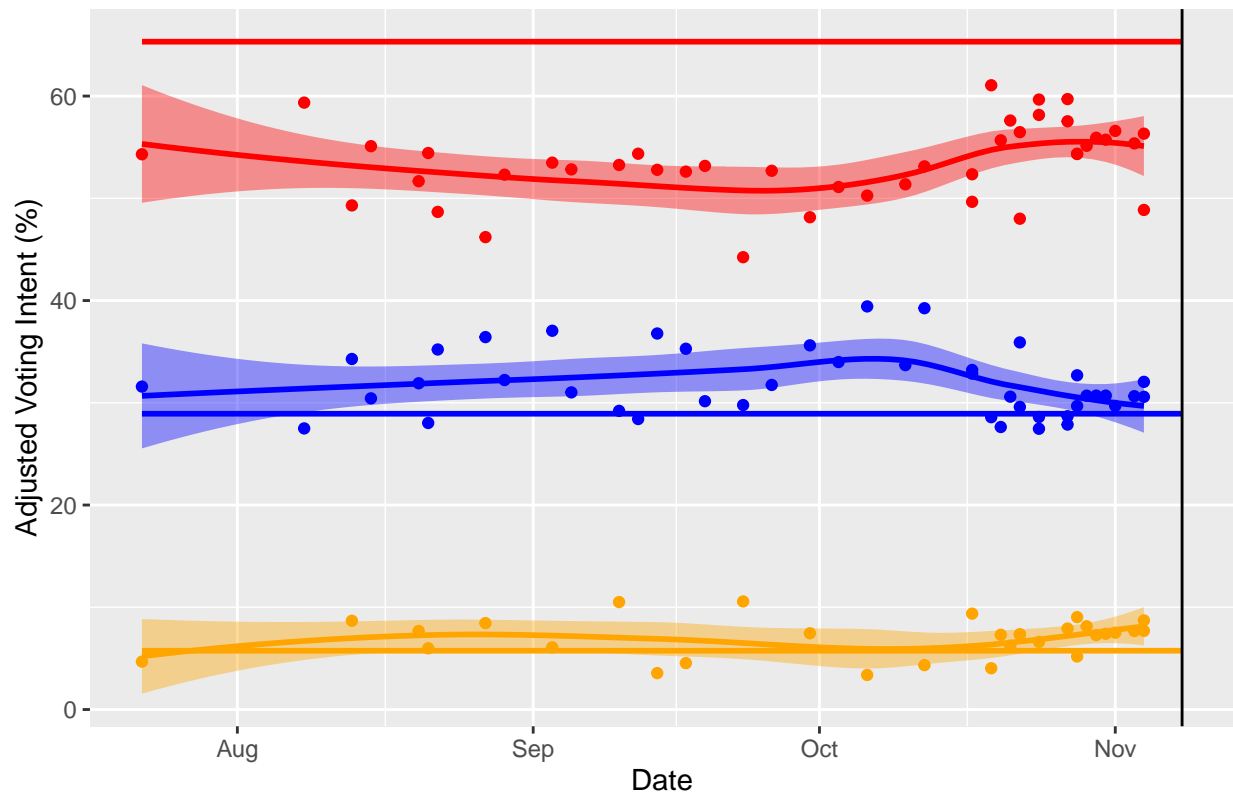
Ohio – Adjusted Polls



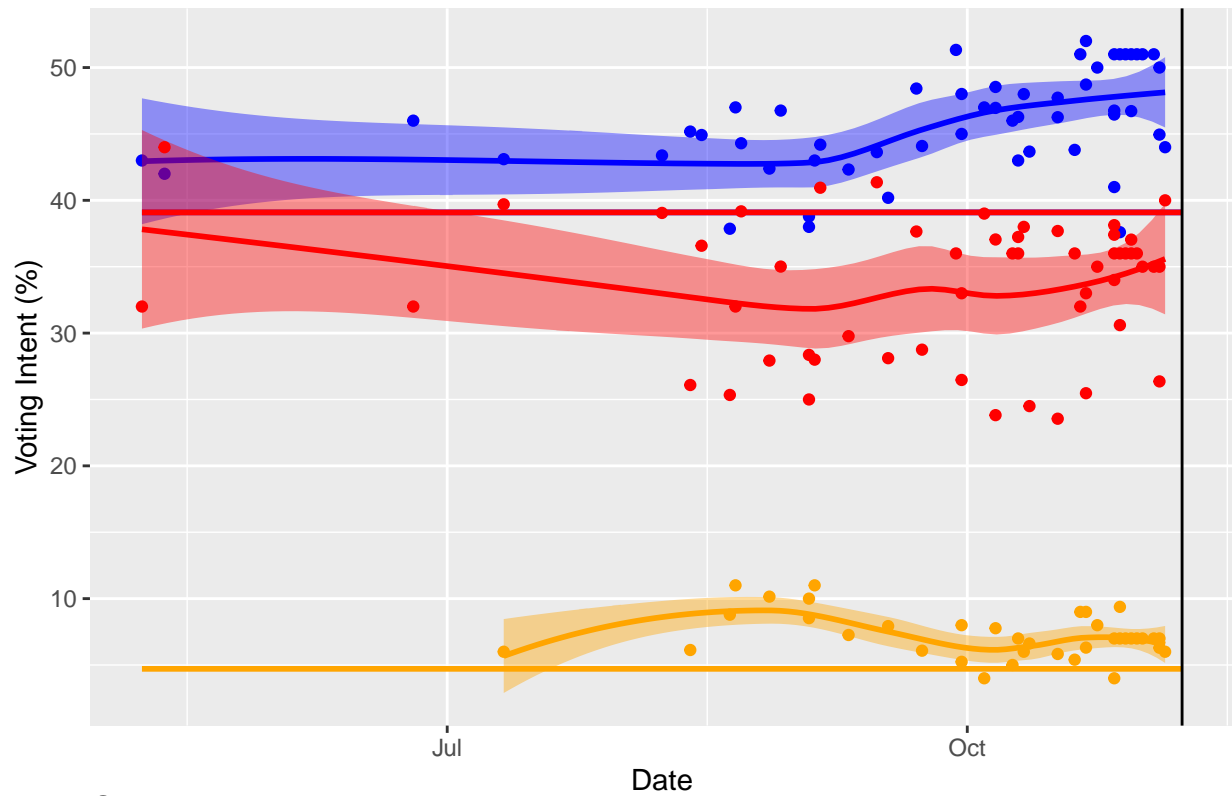
Oklahoma – Raw Polls



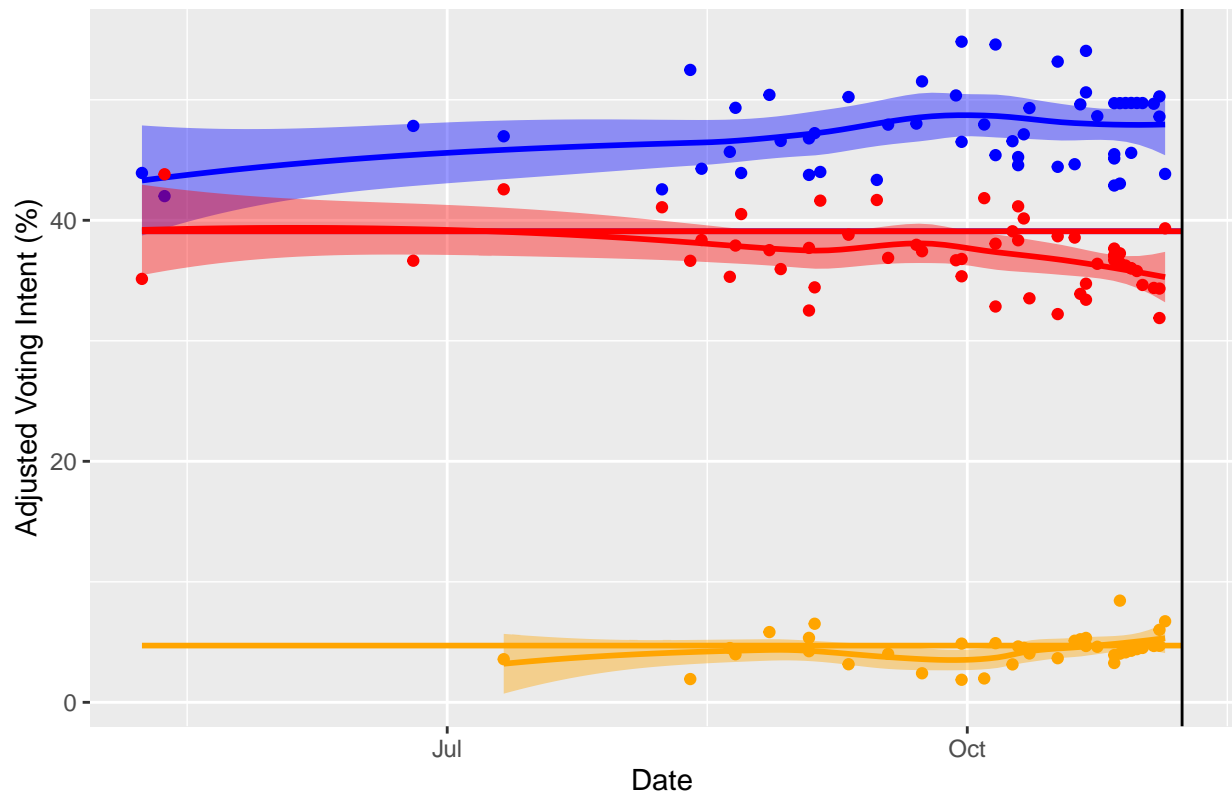
Oklahoma – Adjusted Polls



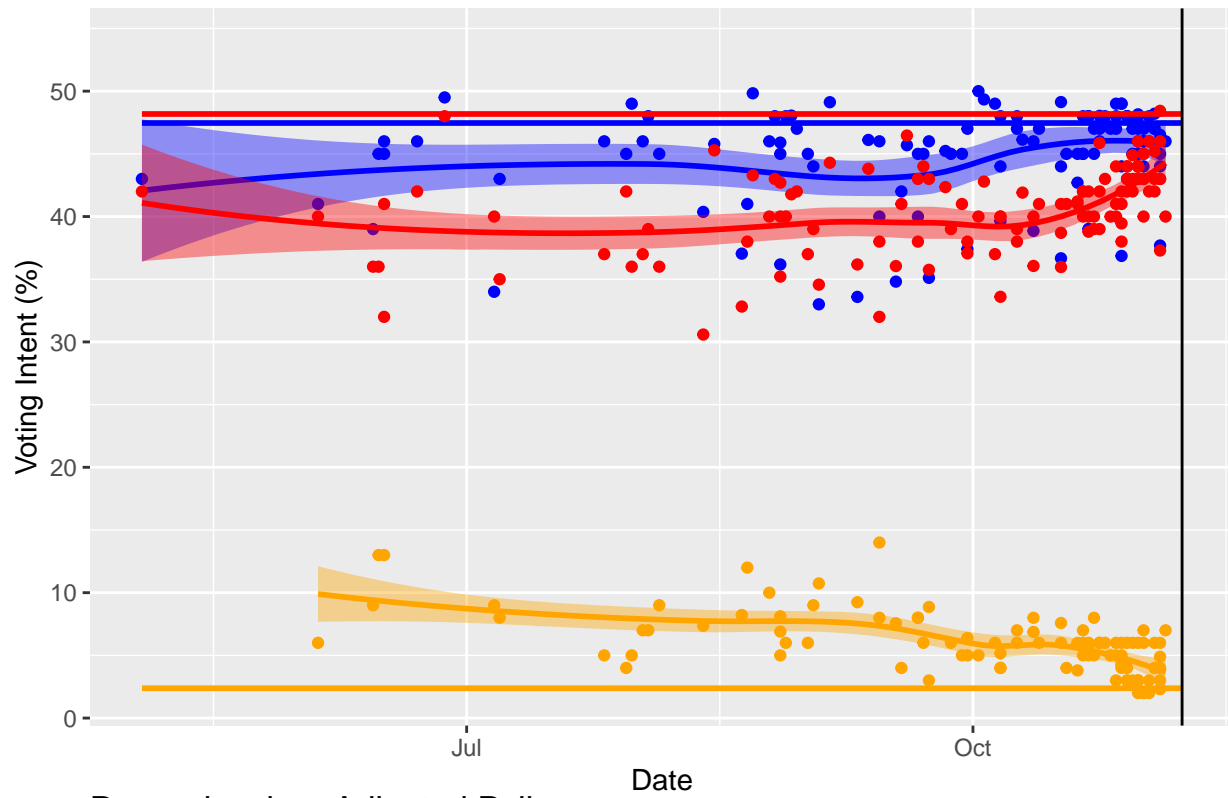
Oregon – Raw Polls



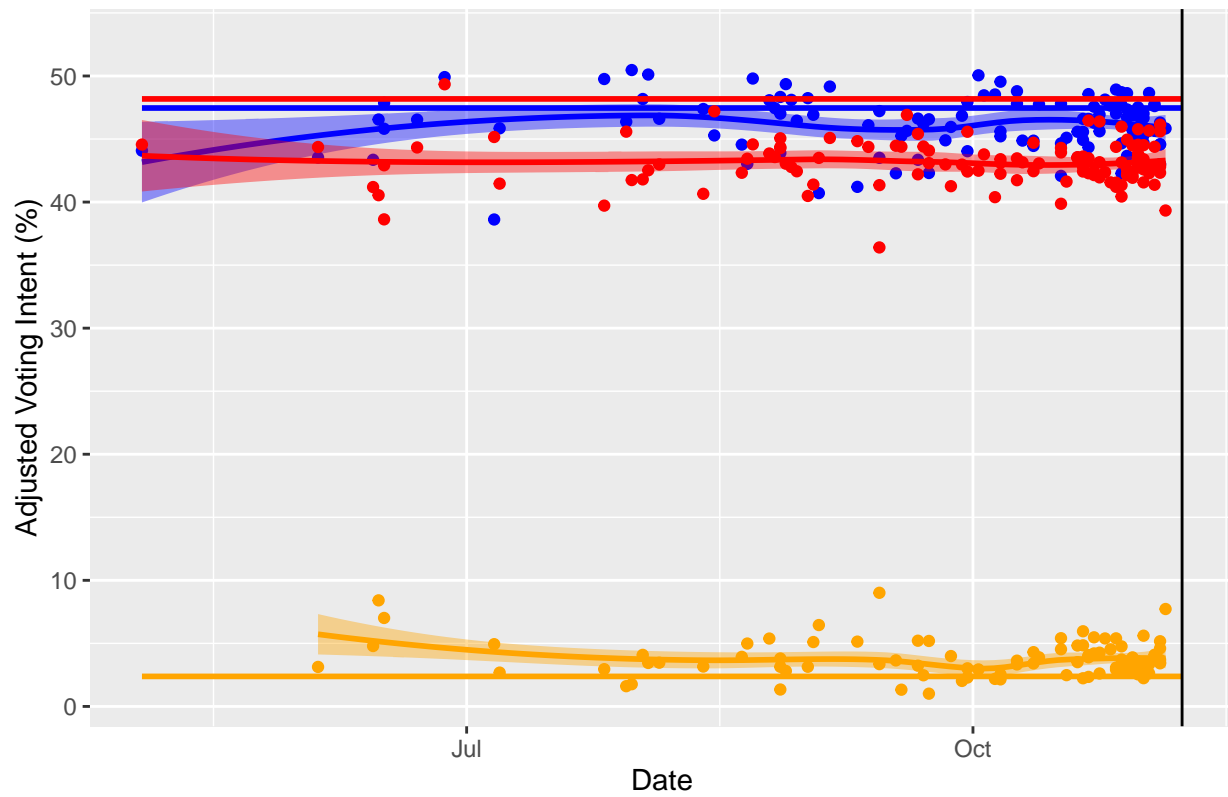
Oregon – Adjusted Polls

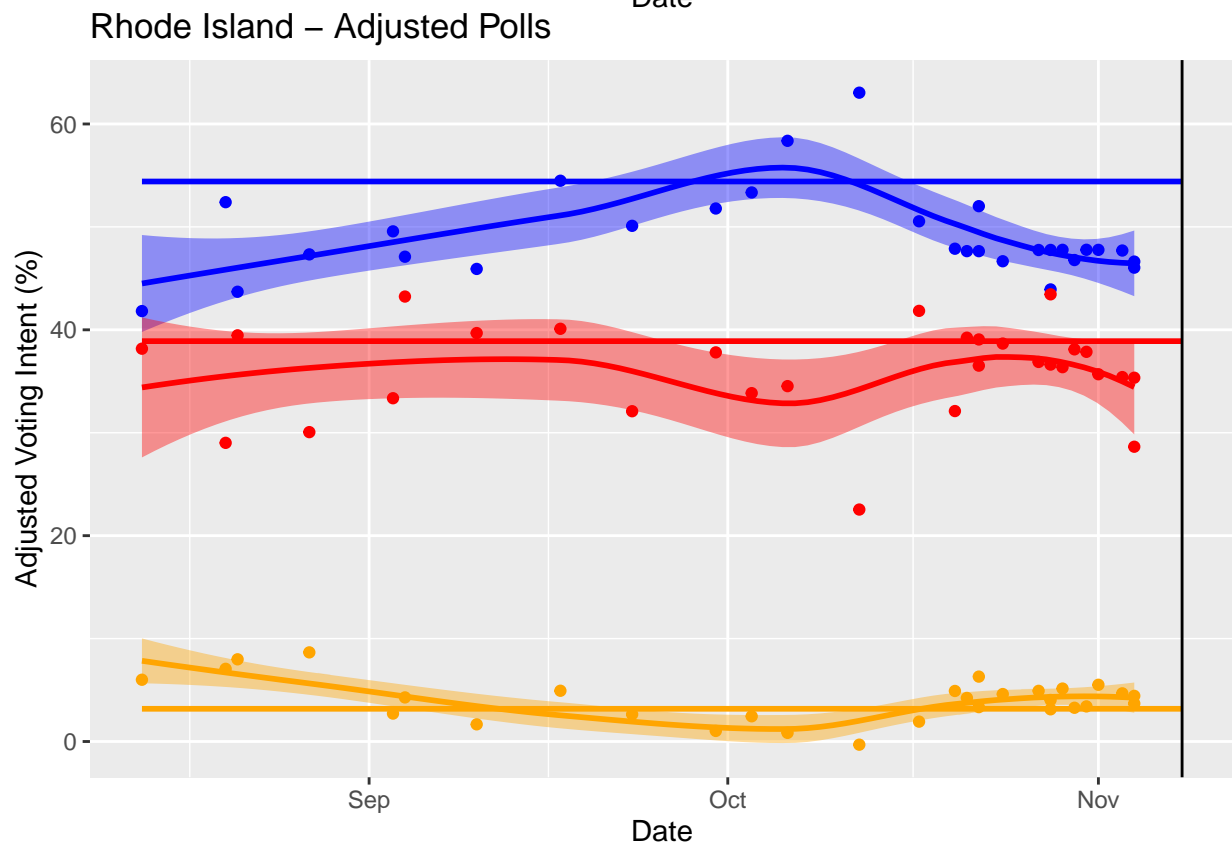
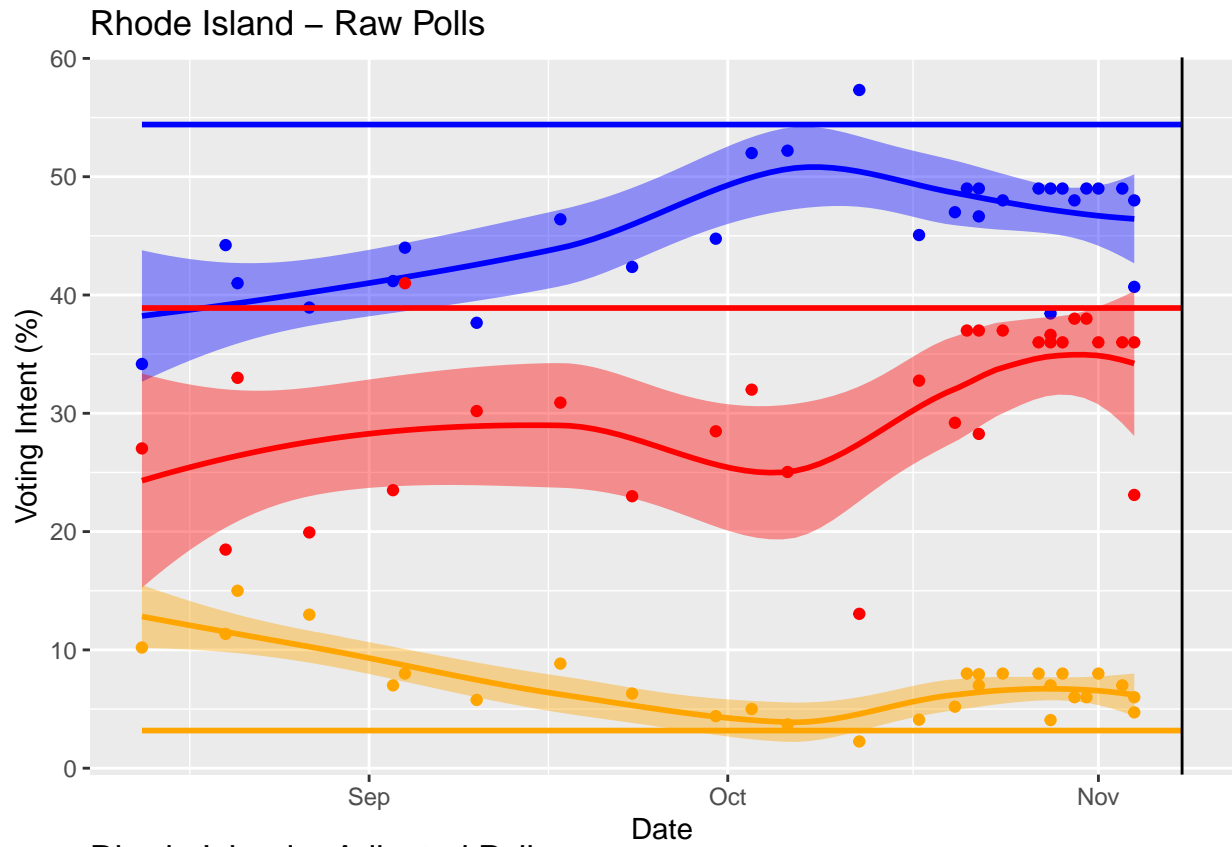


Pennsylvania – Raw Polls

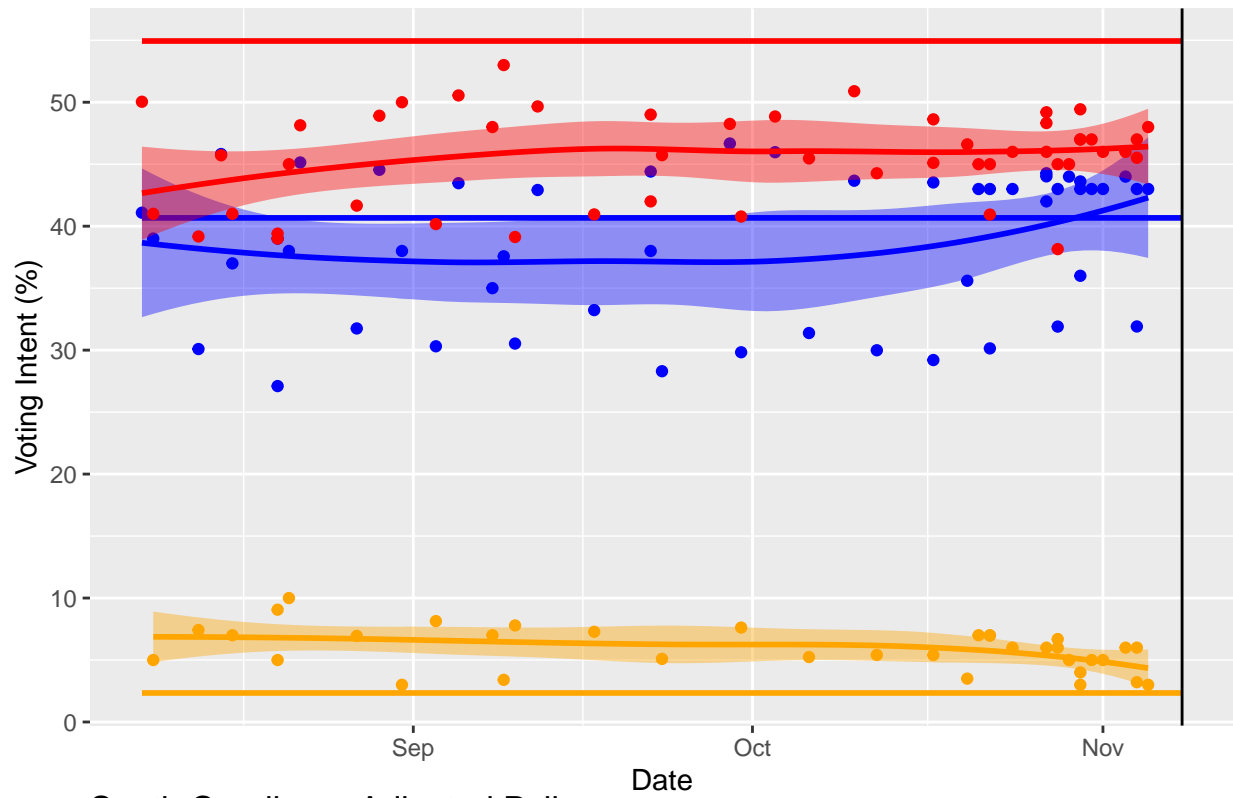


Pennsylvania – Adjusted Polls

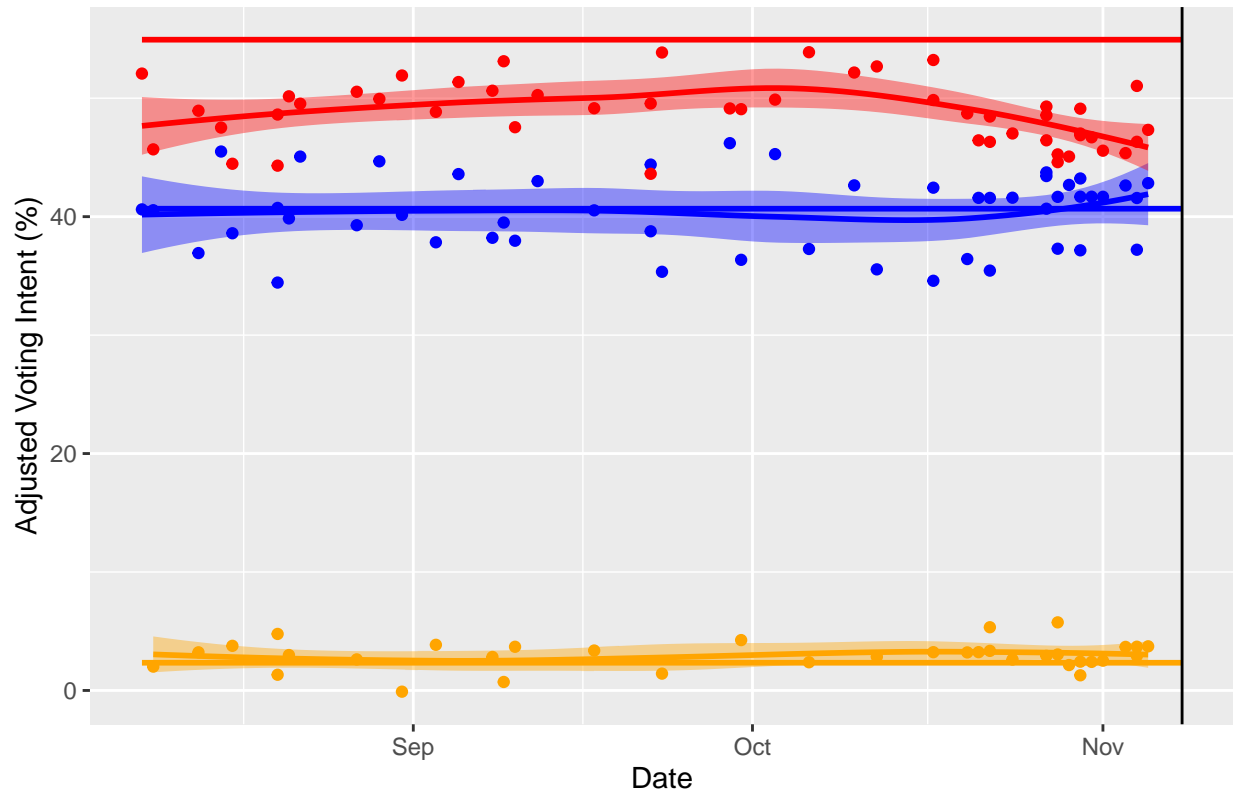




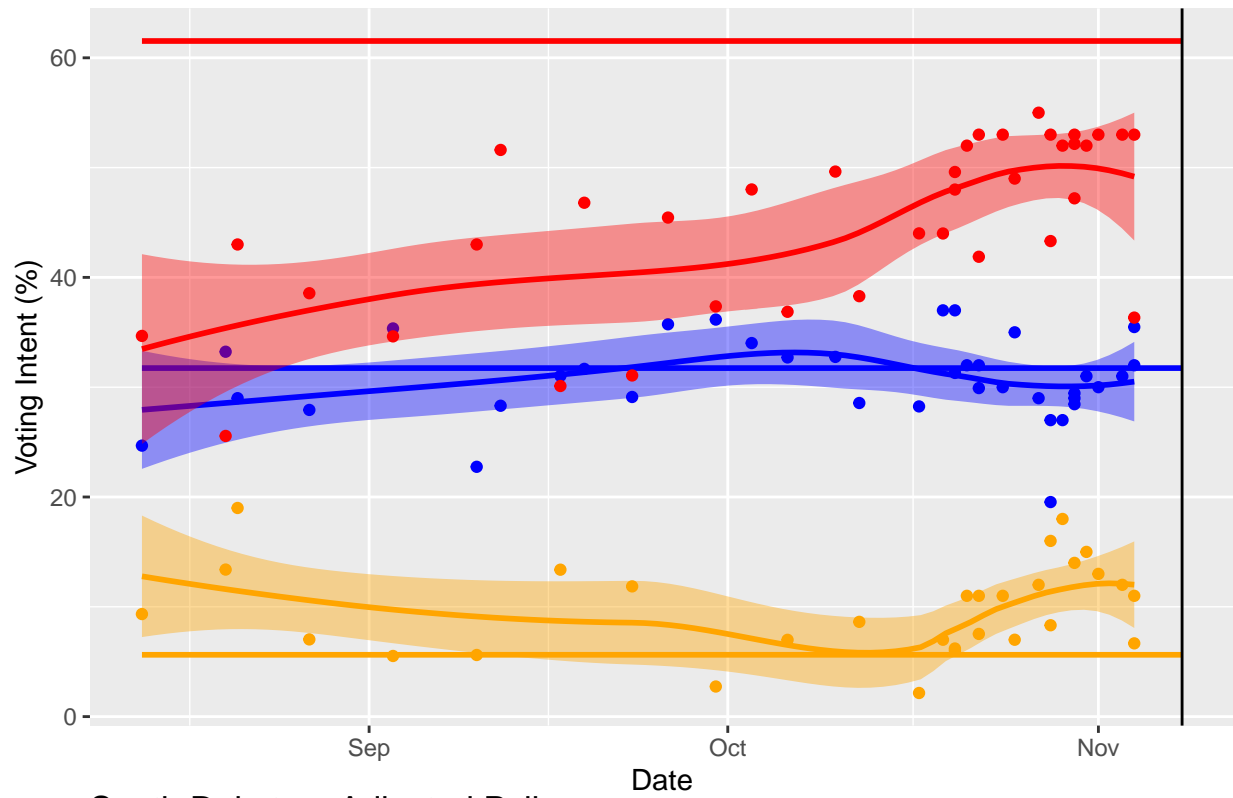
South Carolina – Raw Polls



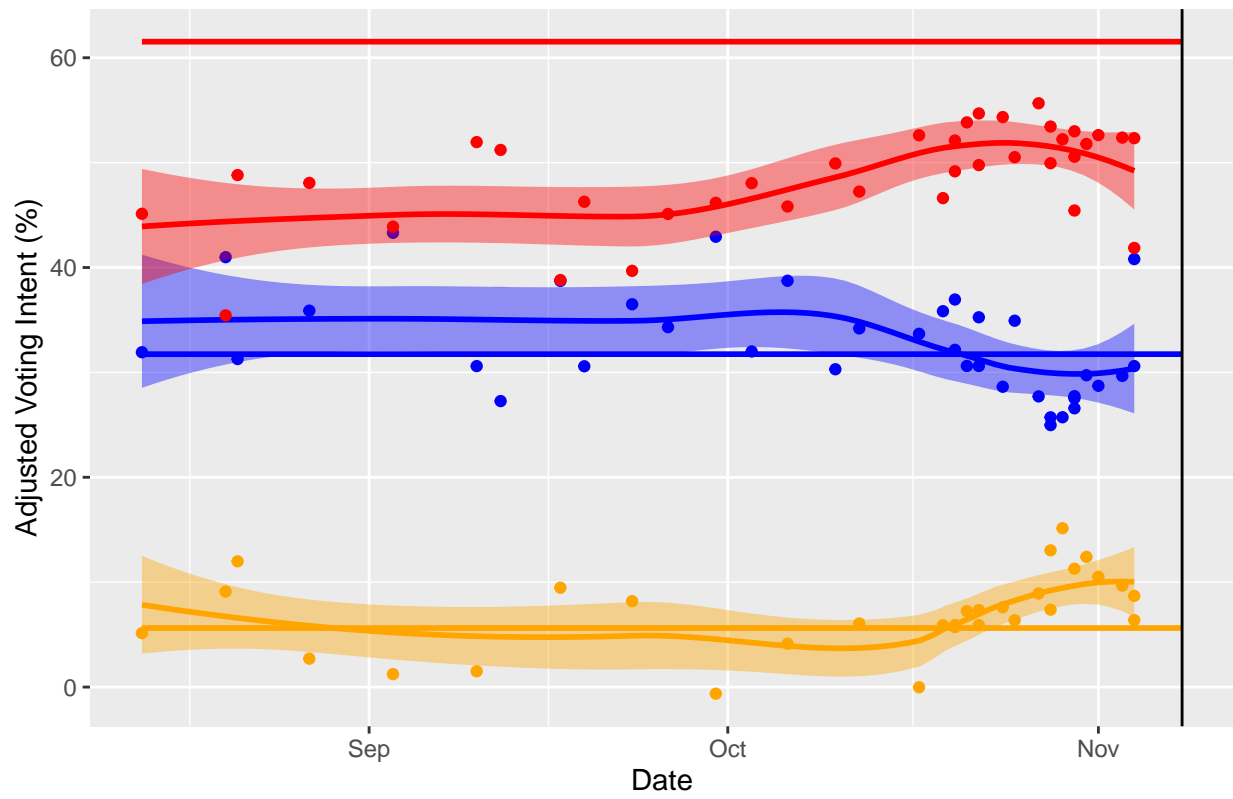
South Carolina – Adjusted Polls



South Dakota – Raw Polls

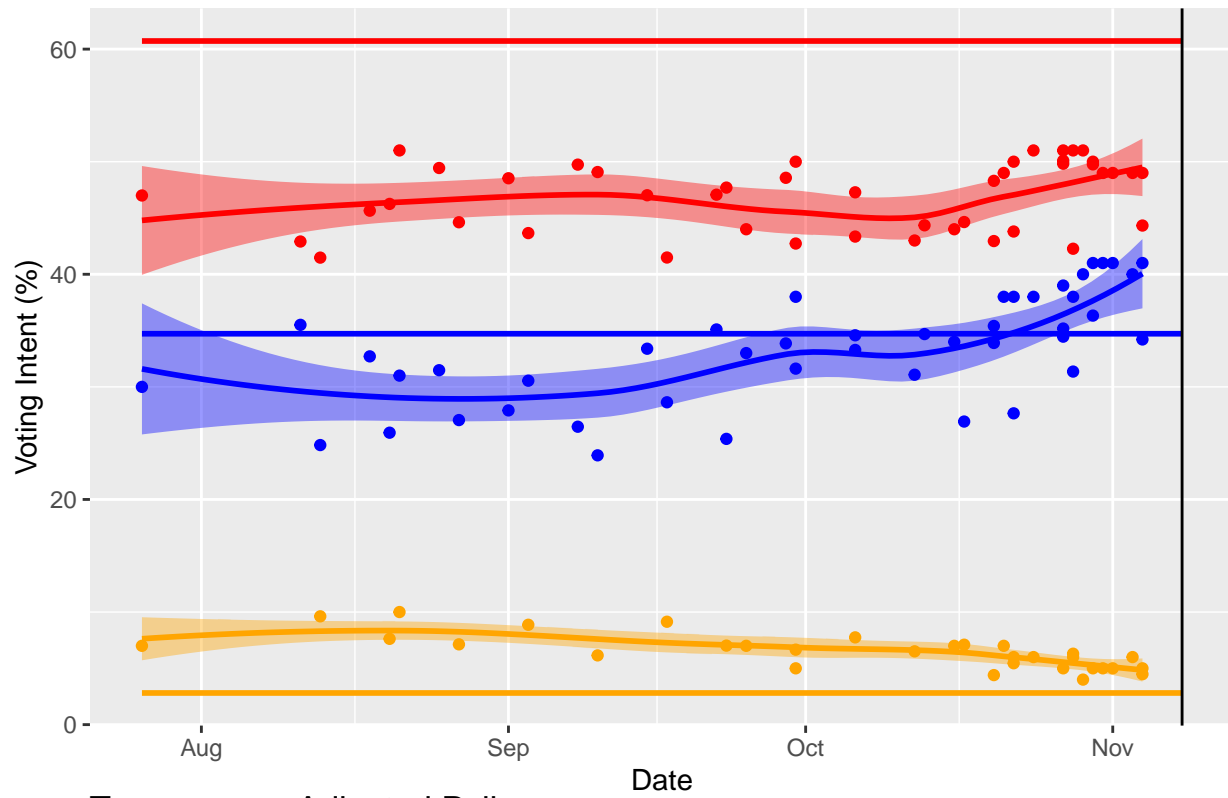


South Dakota – Adjusted Polls

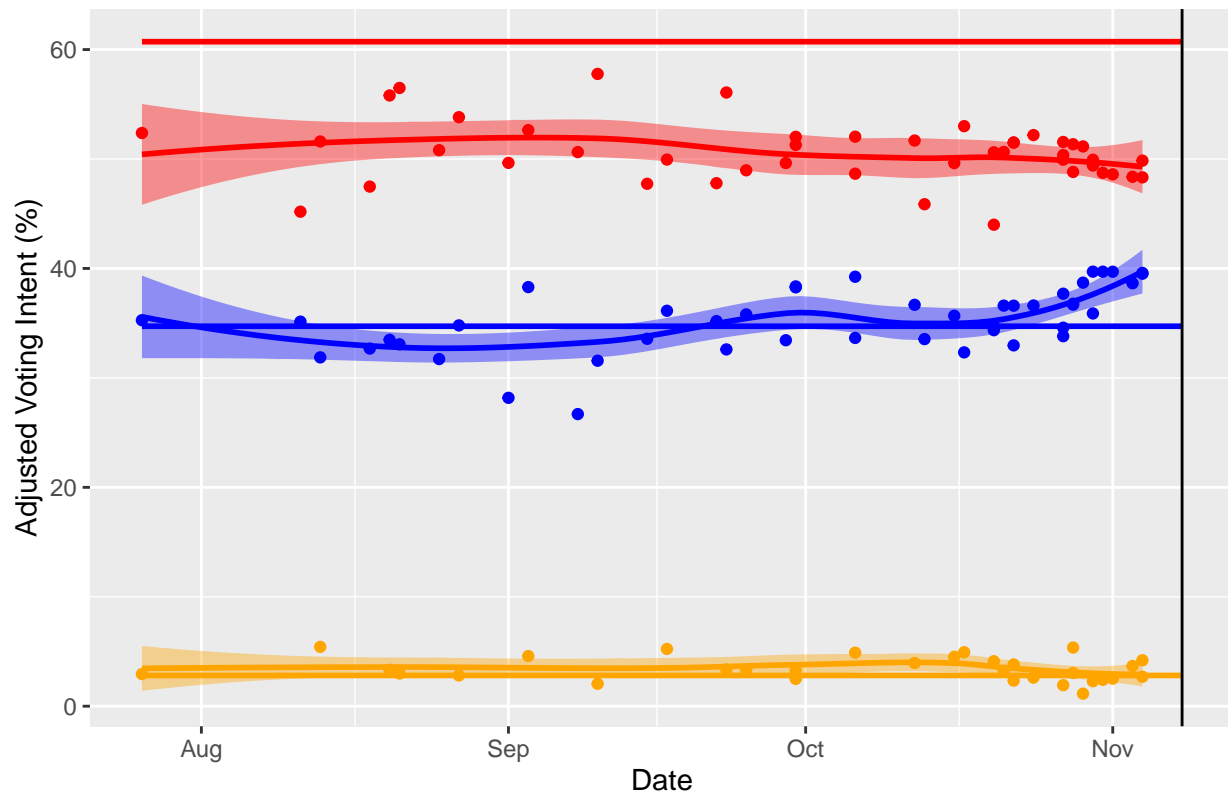




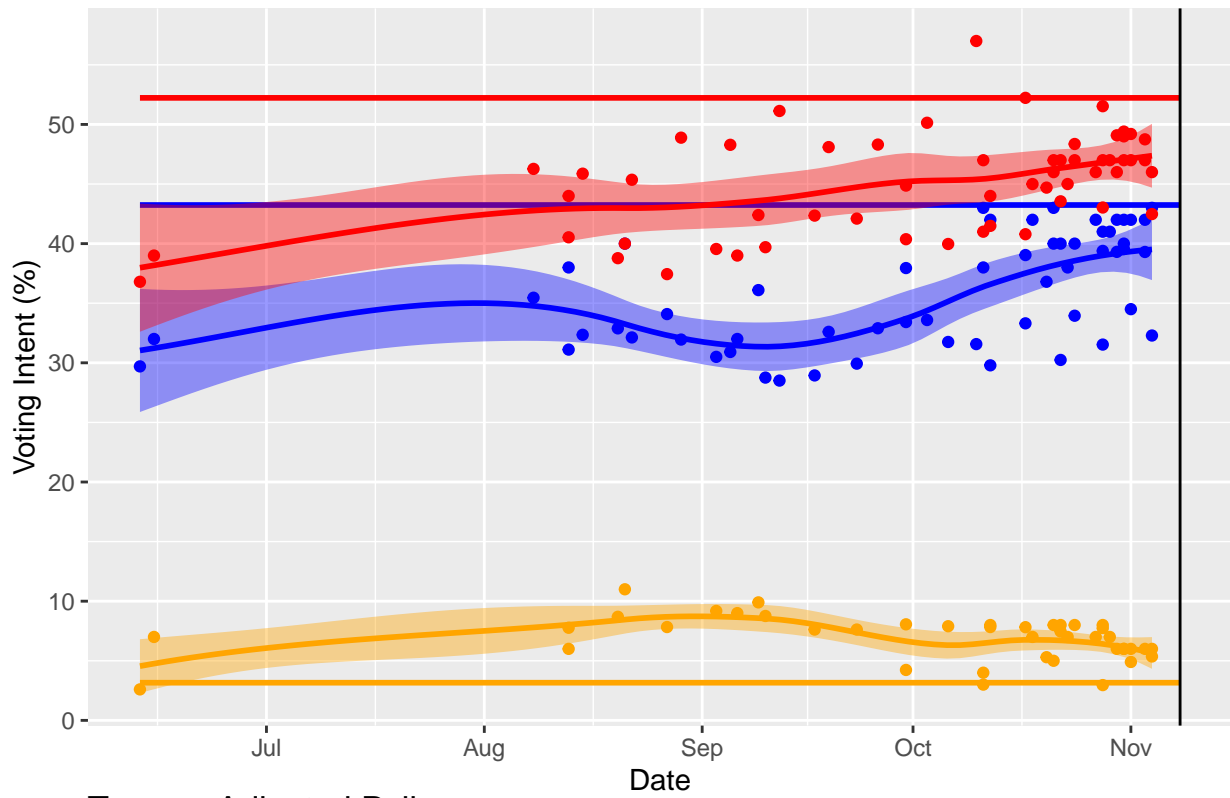
Tennessee – Raw Polls



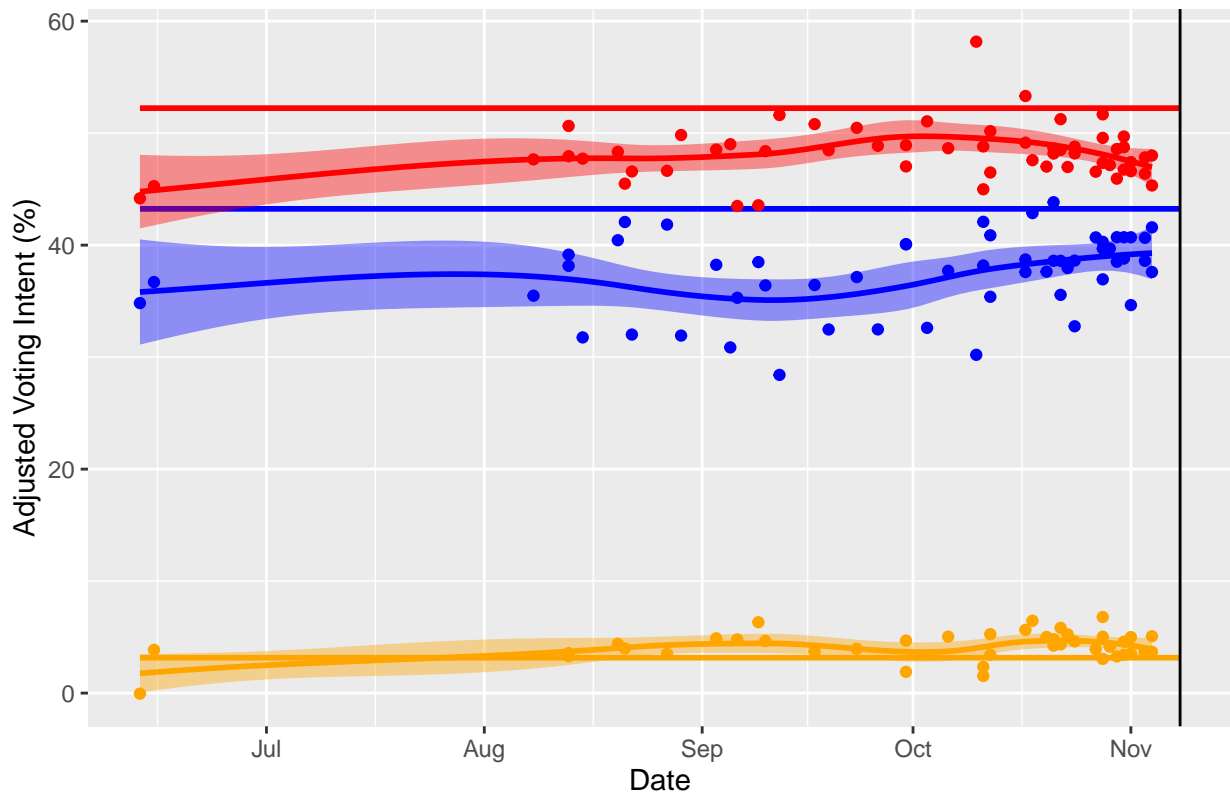
Tennessee – Adjusted Polls

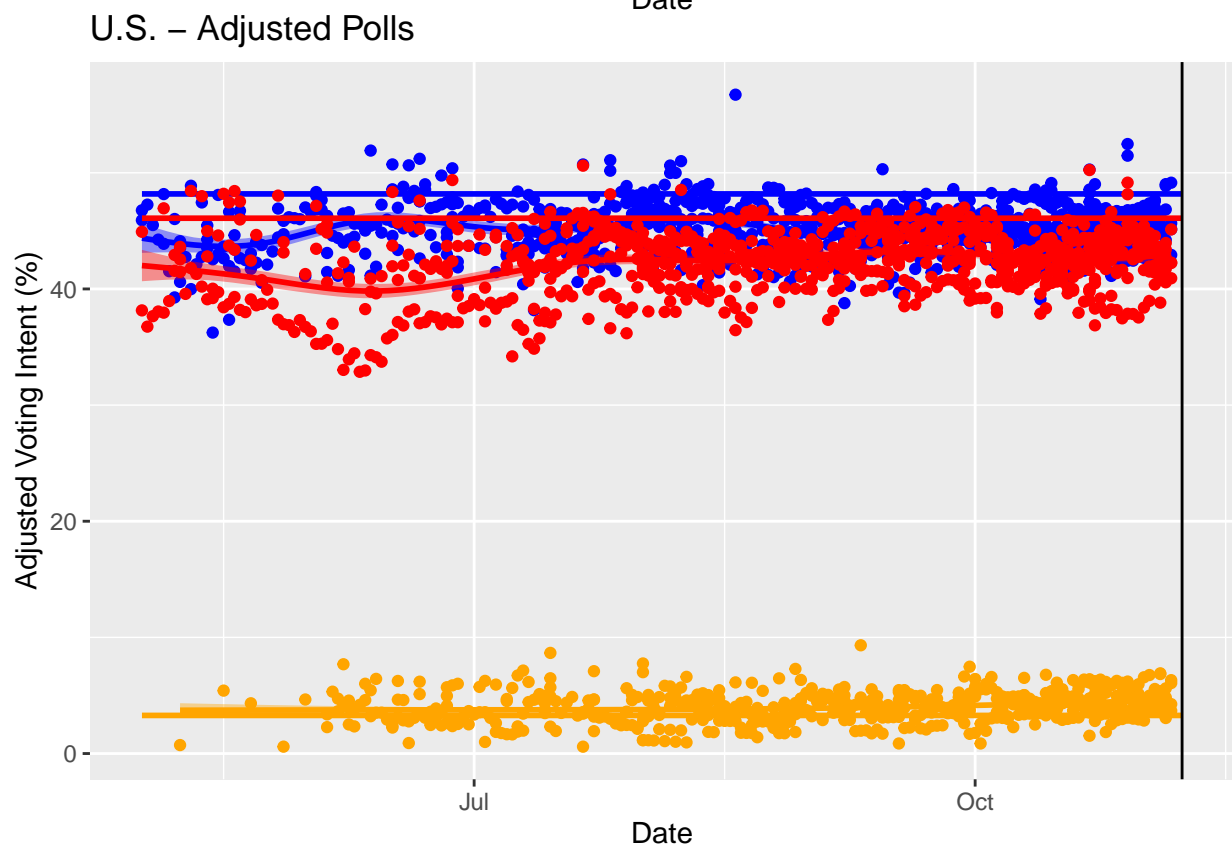
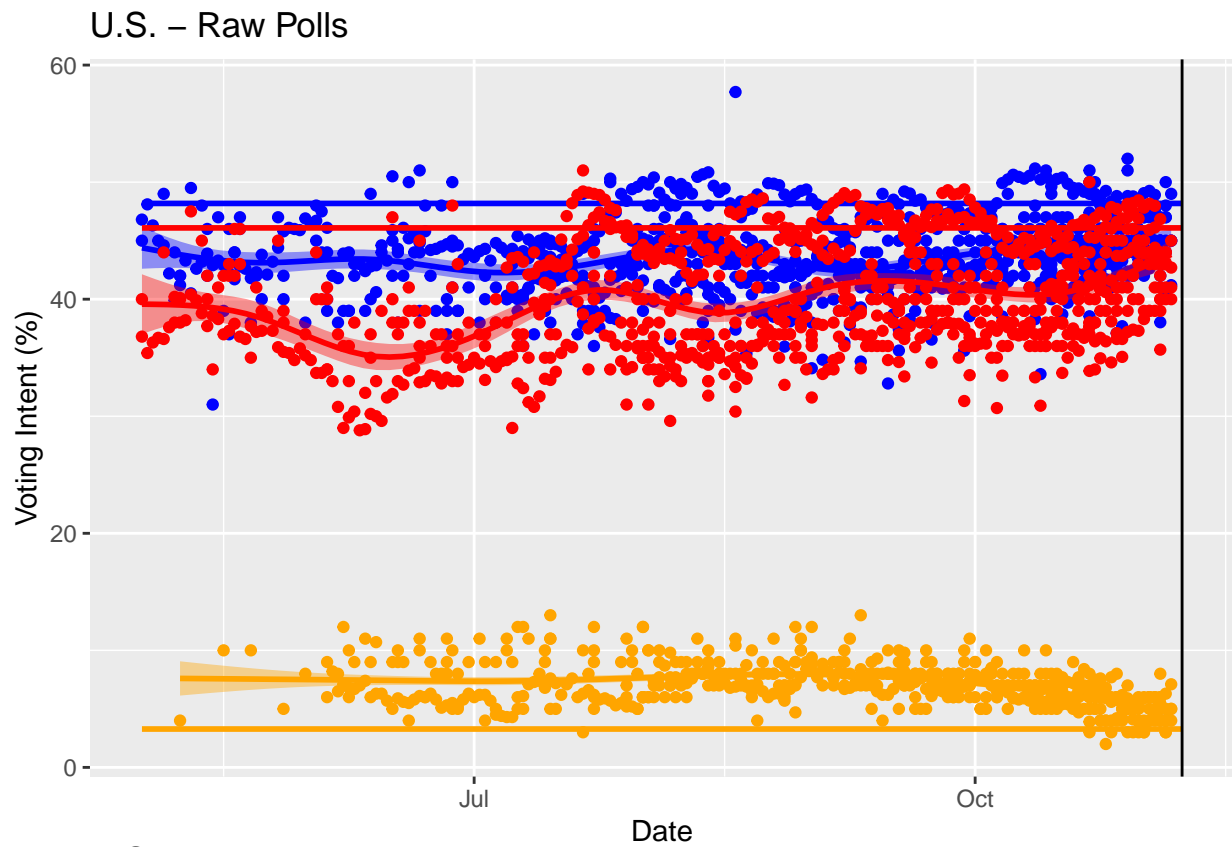


Texas – Raw Polls

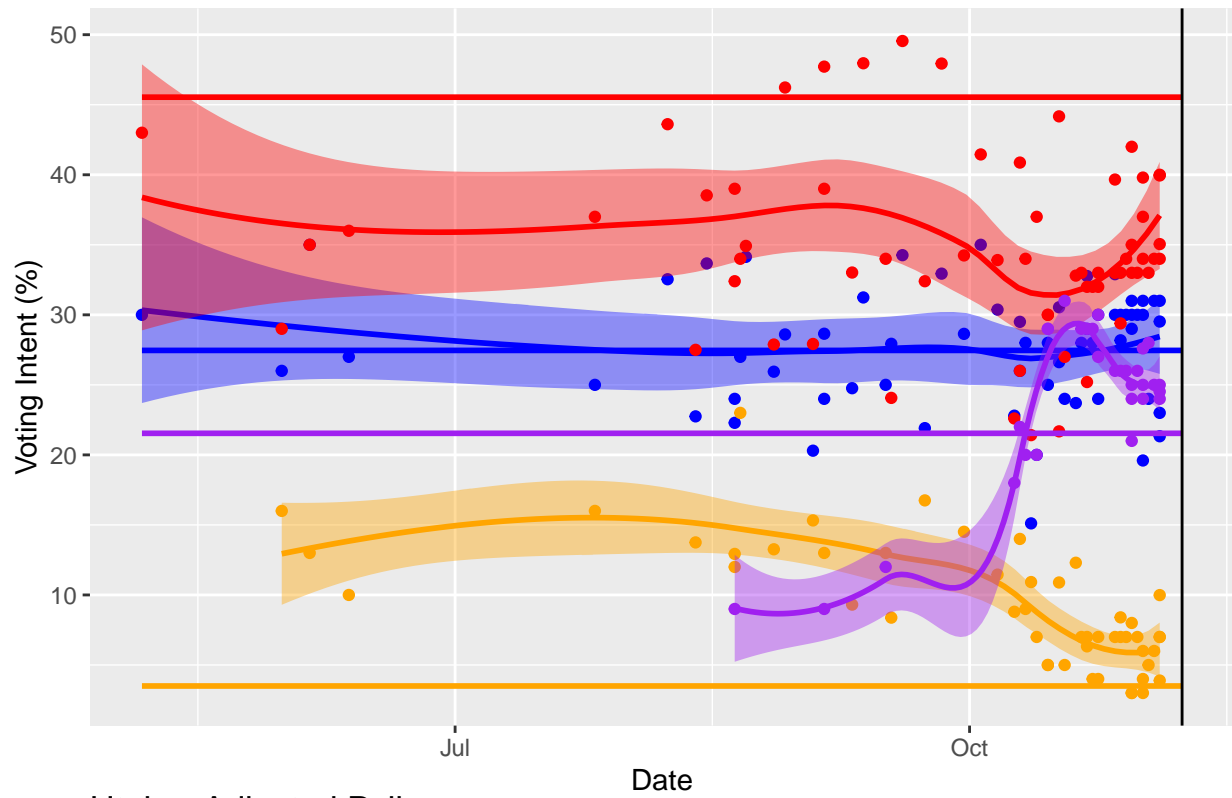


Texas – Adjusted Polls

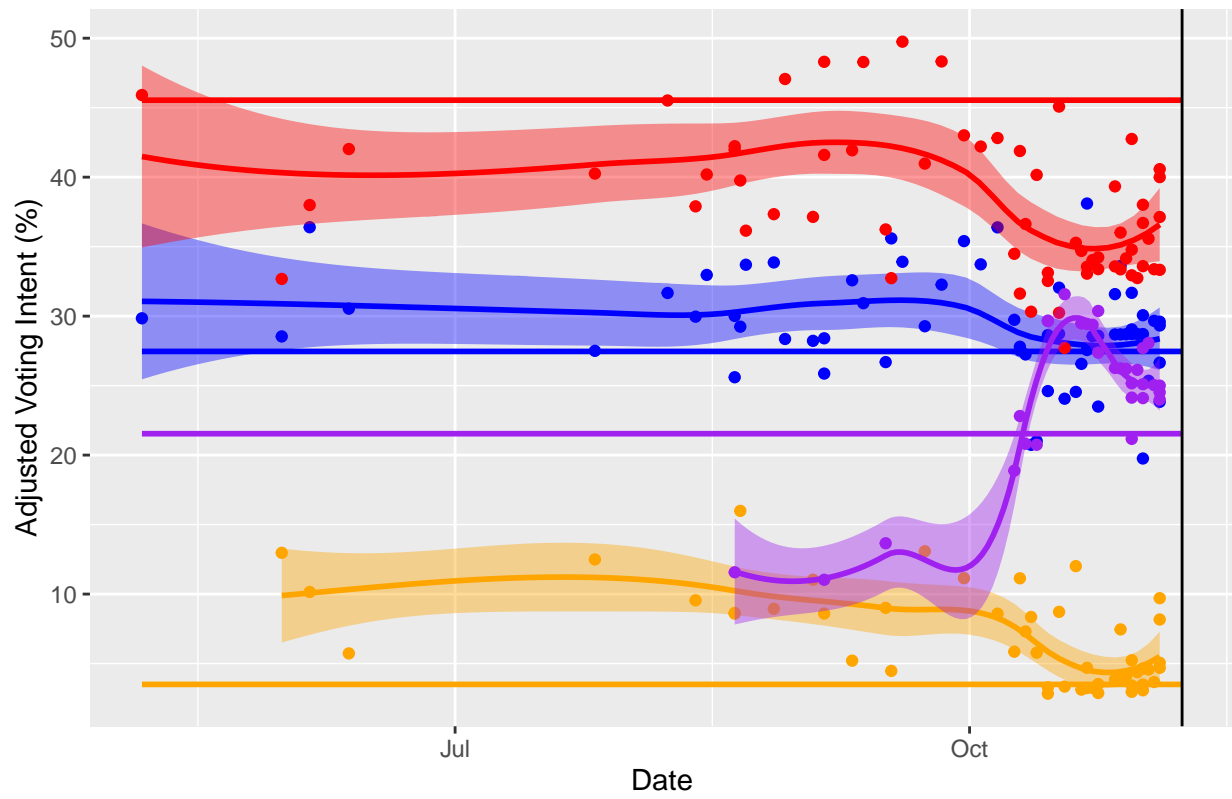




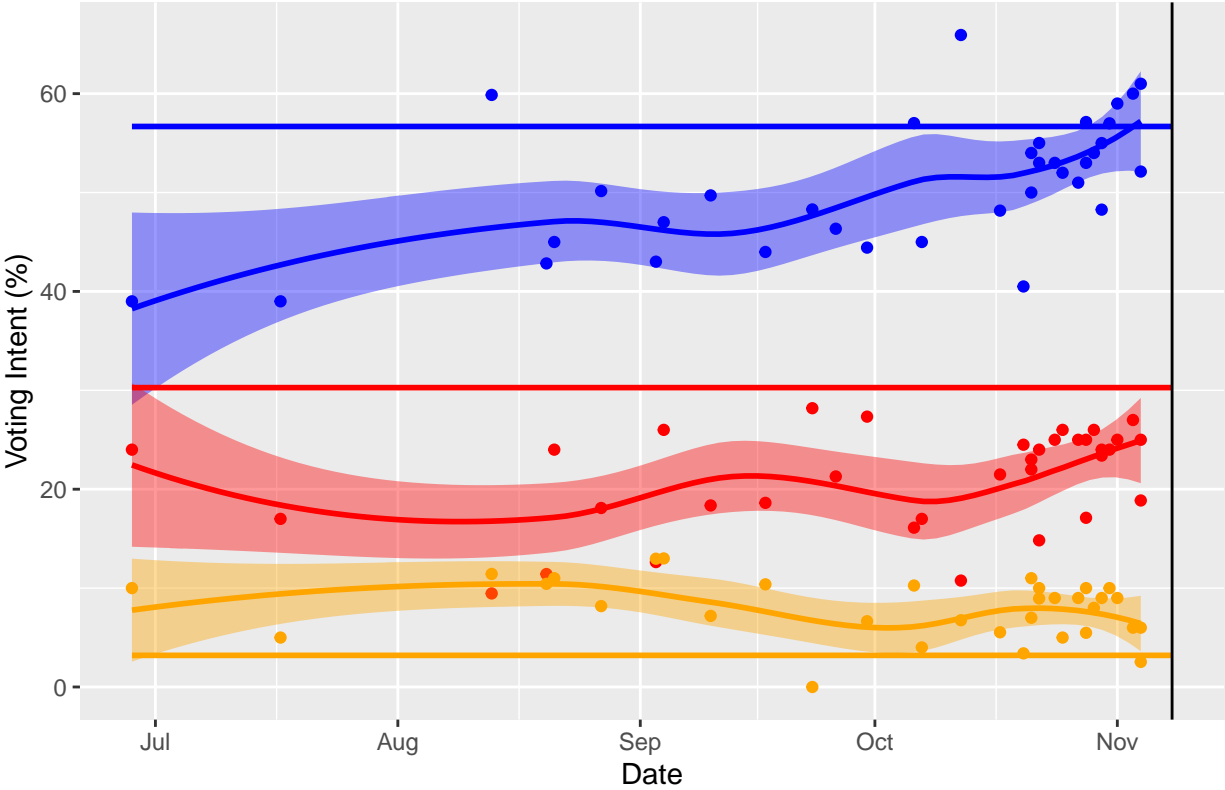
Utah – Raw Polls



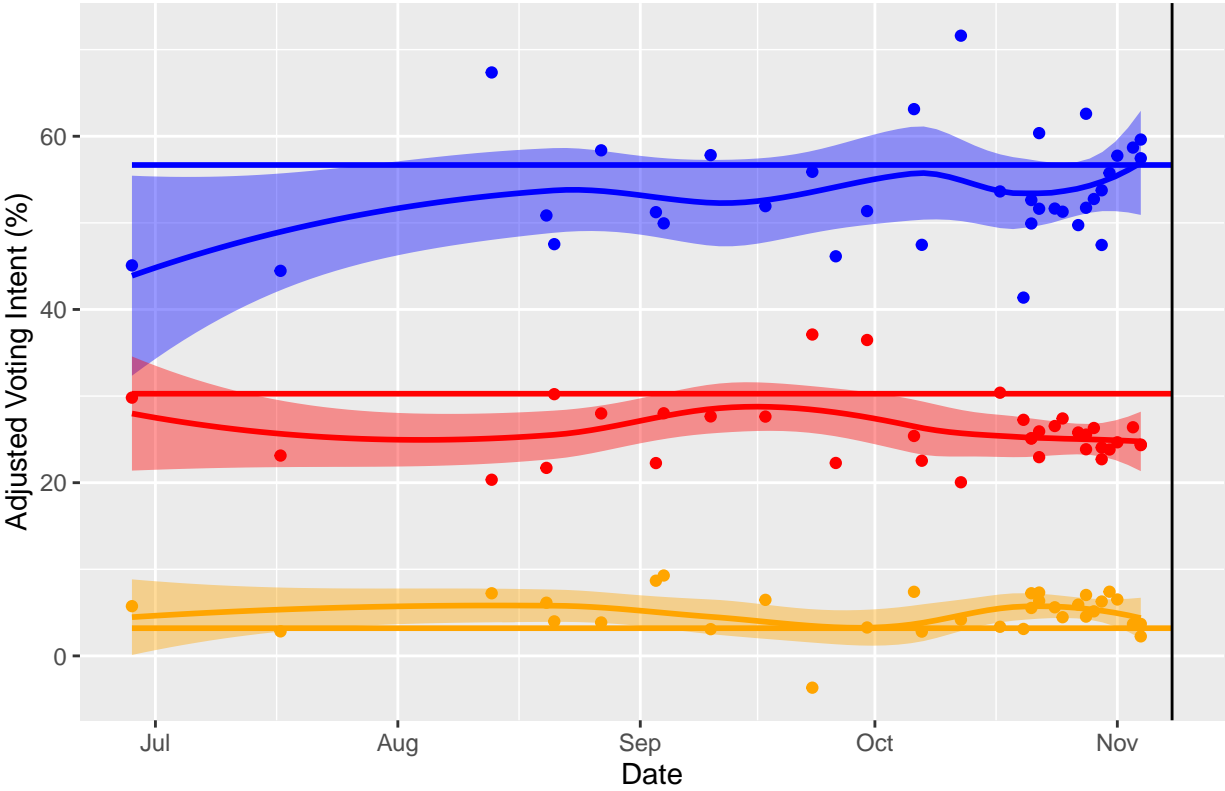
Utah – Adjusted Polls



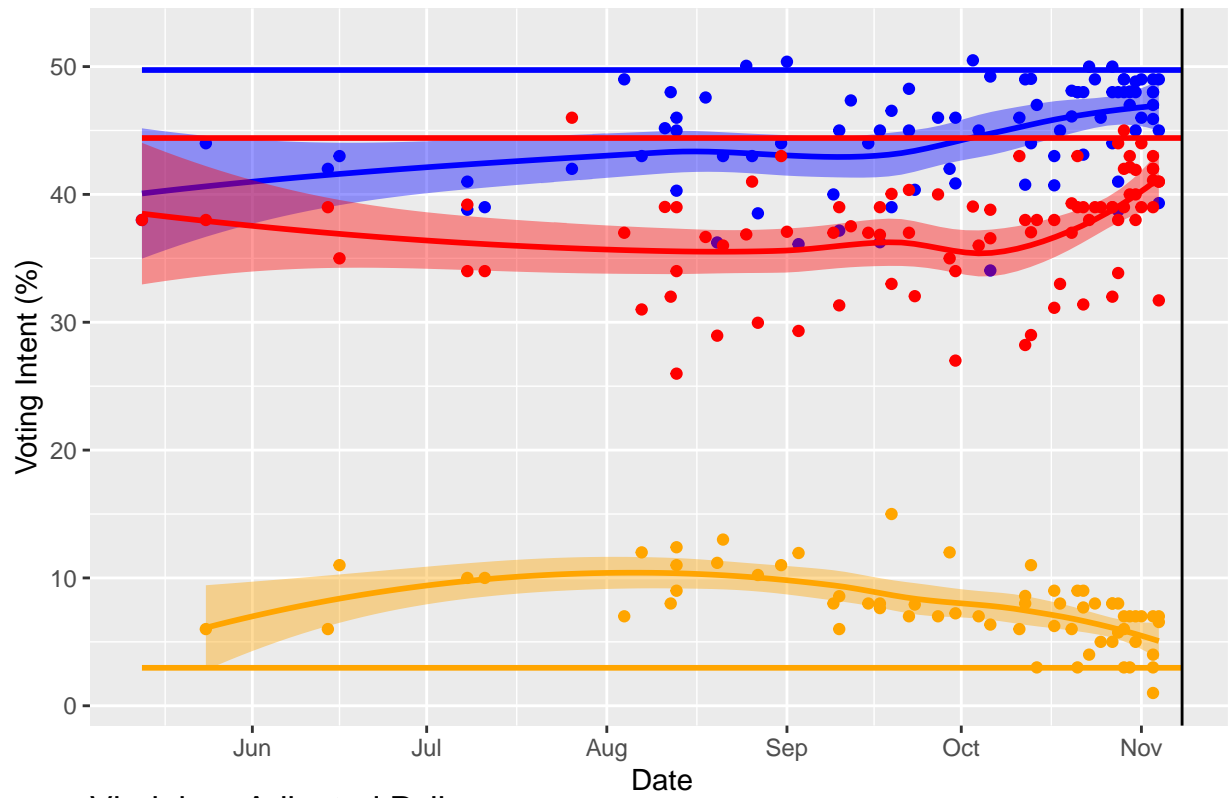
Vermont – Raw Polls



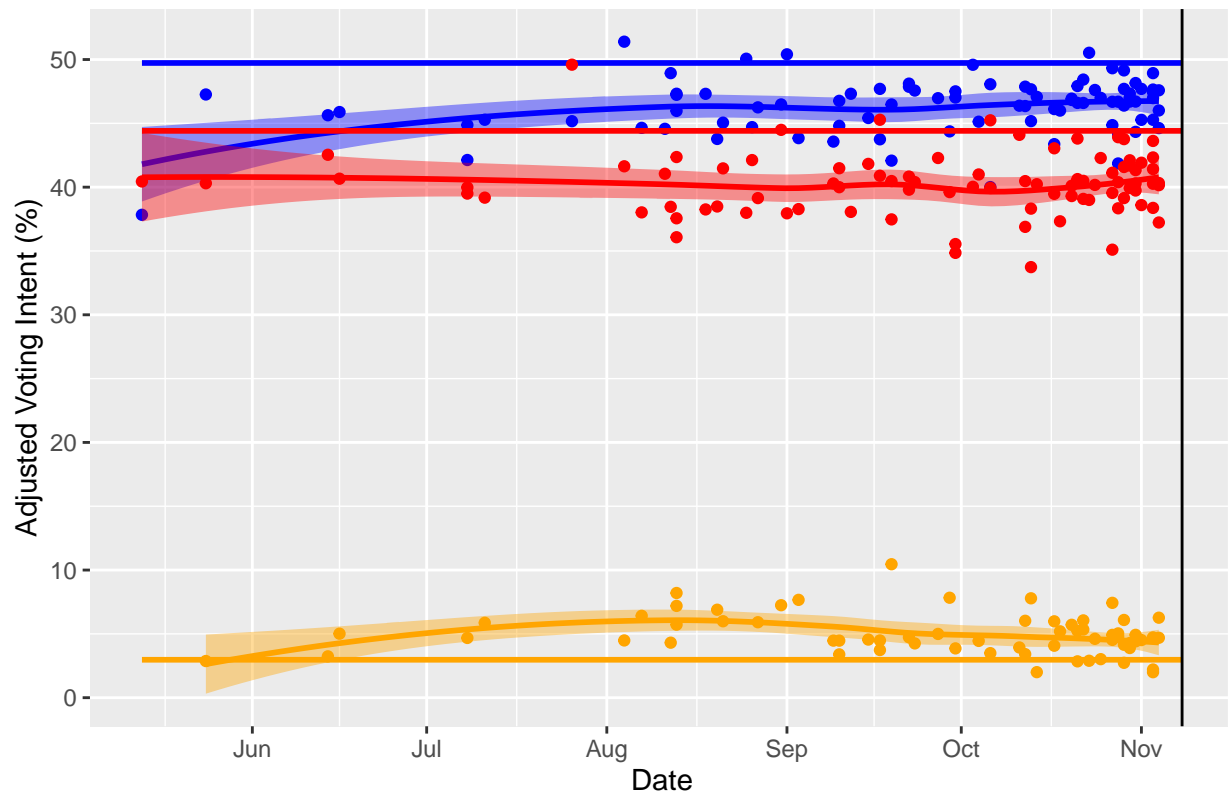
Vermont – Adjusted Polls



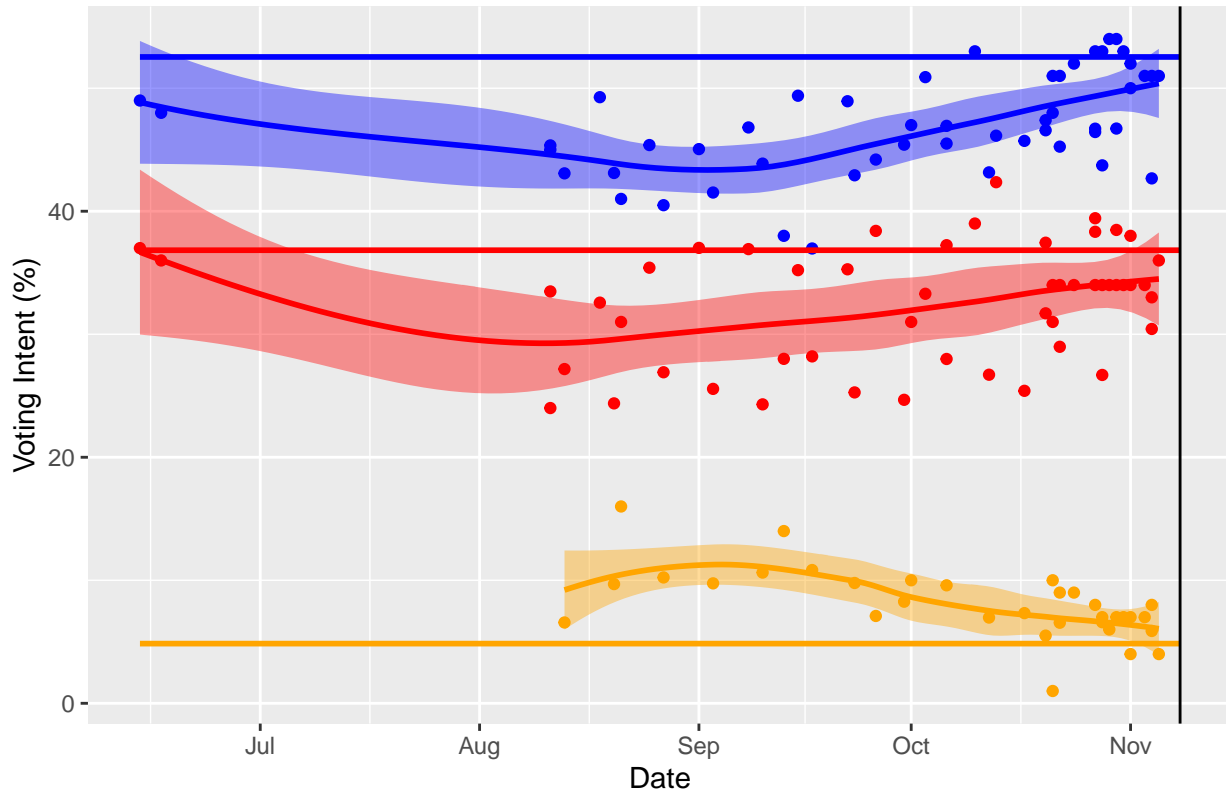
Virginia – Raw Polls



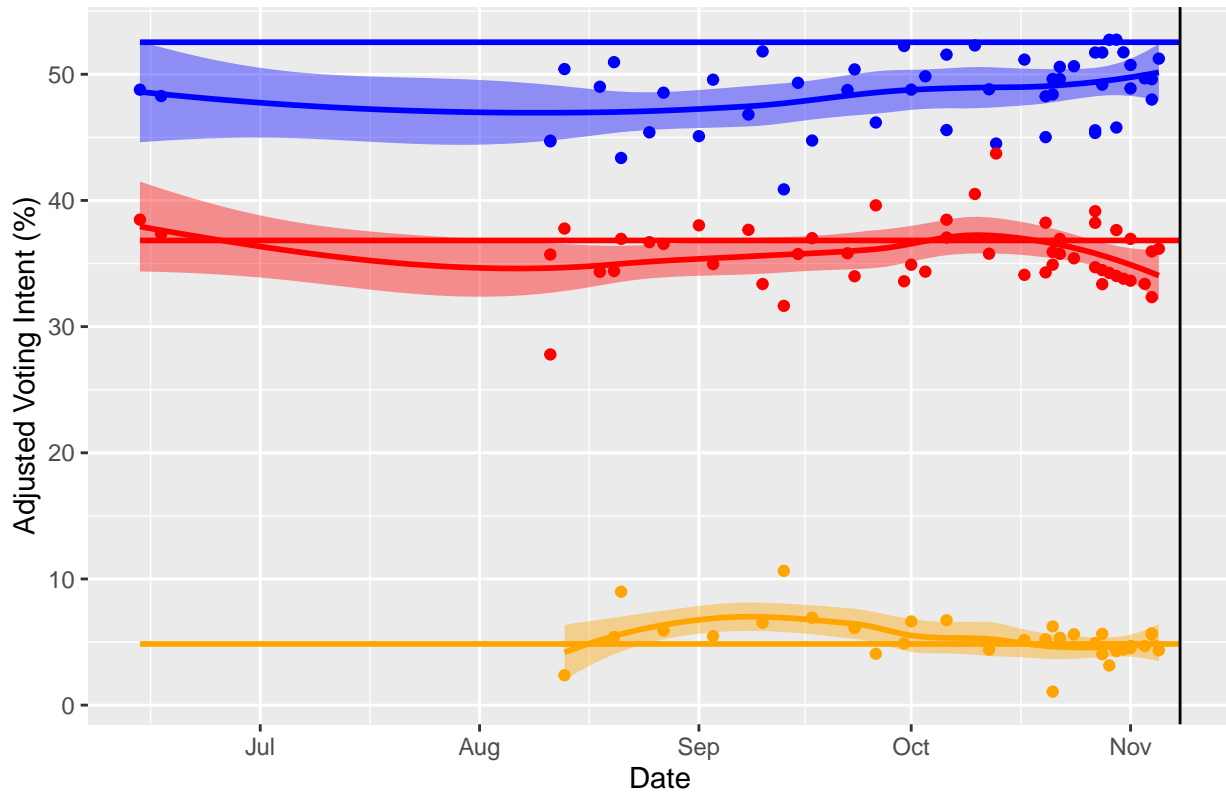
Virginia – Adjusted Polls



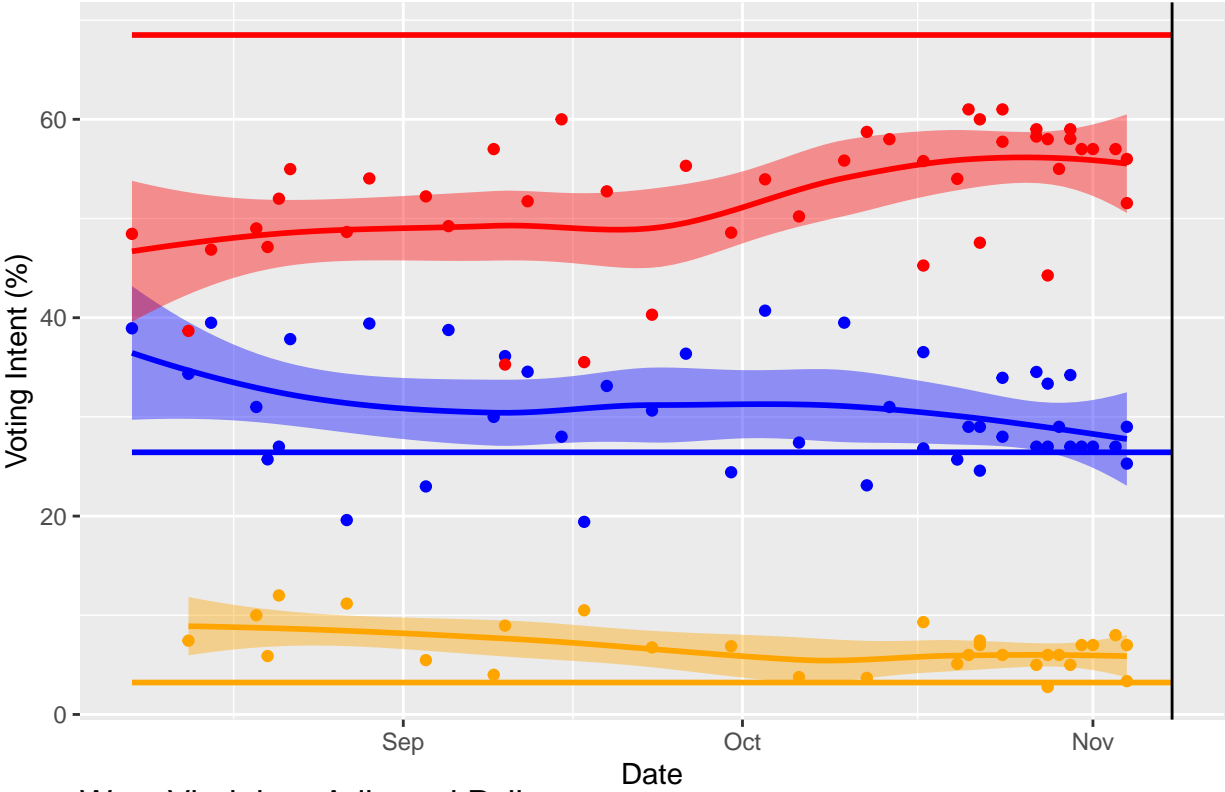
Washington – Raw Polls



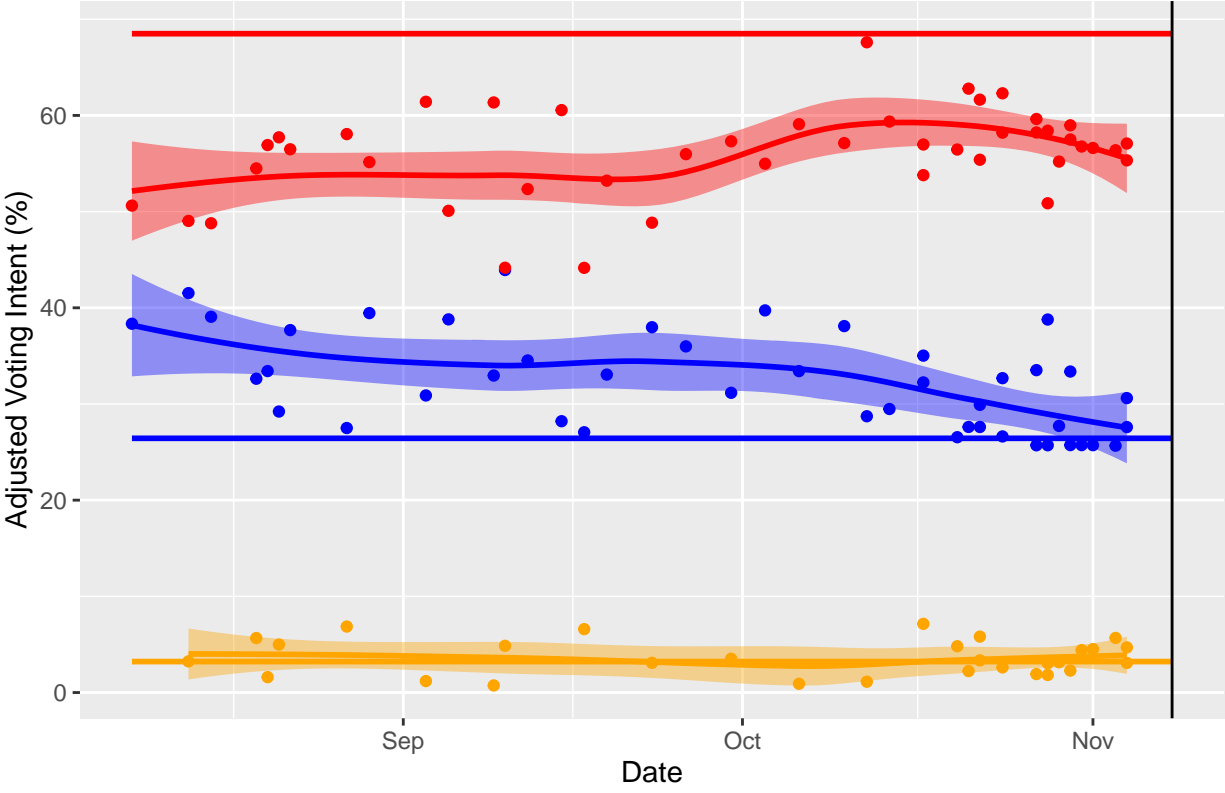
Washington – Adjusted Polls



West Virginia – Raw Polls

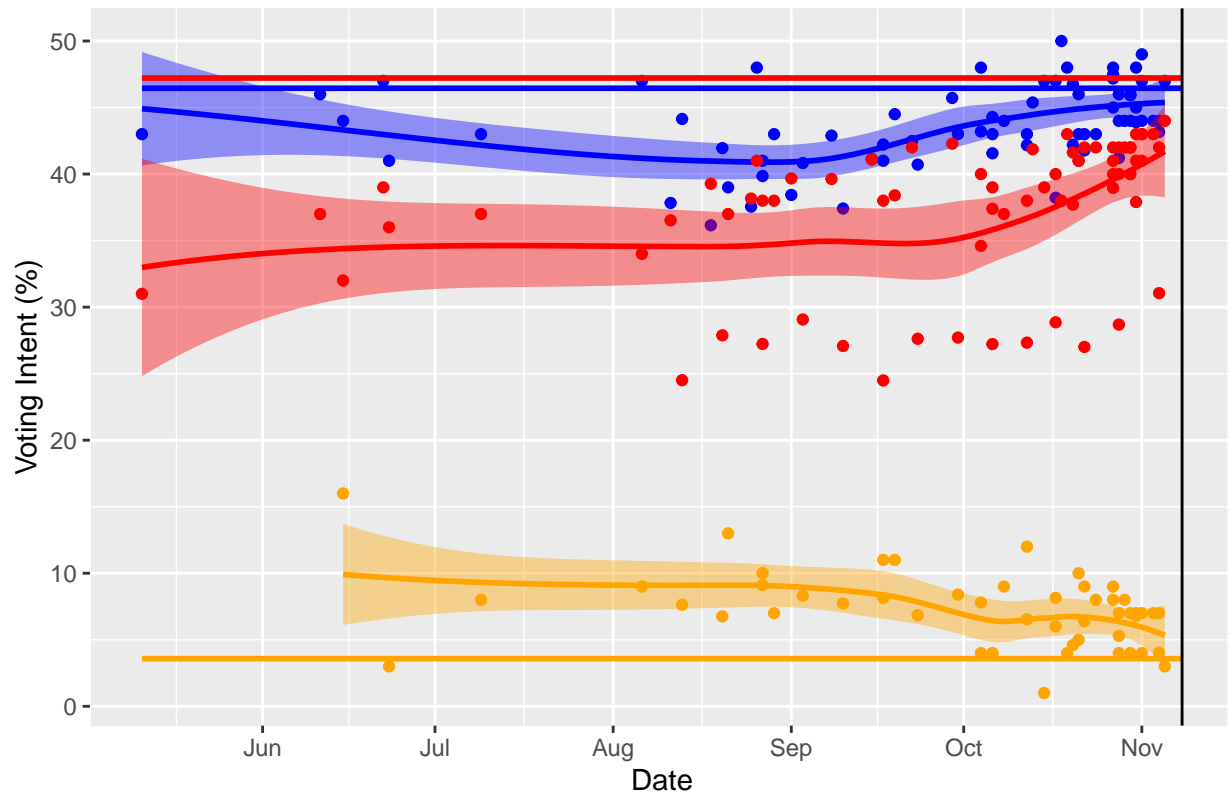


West Virginia – Adjusted Polls

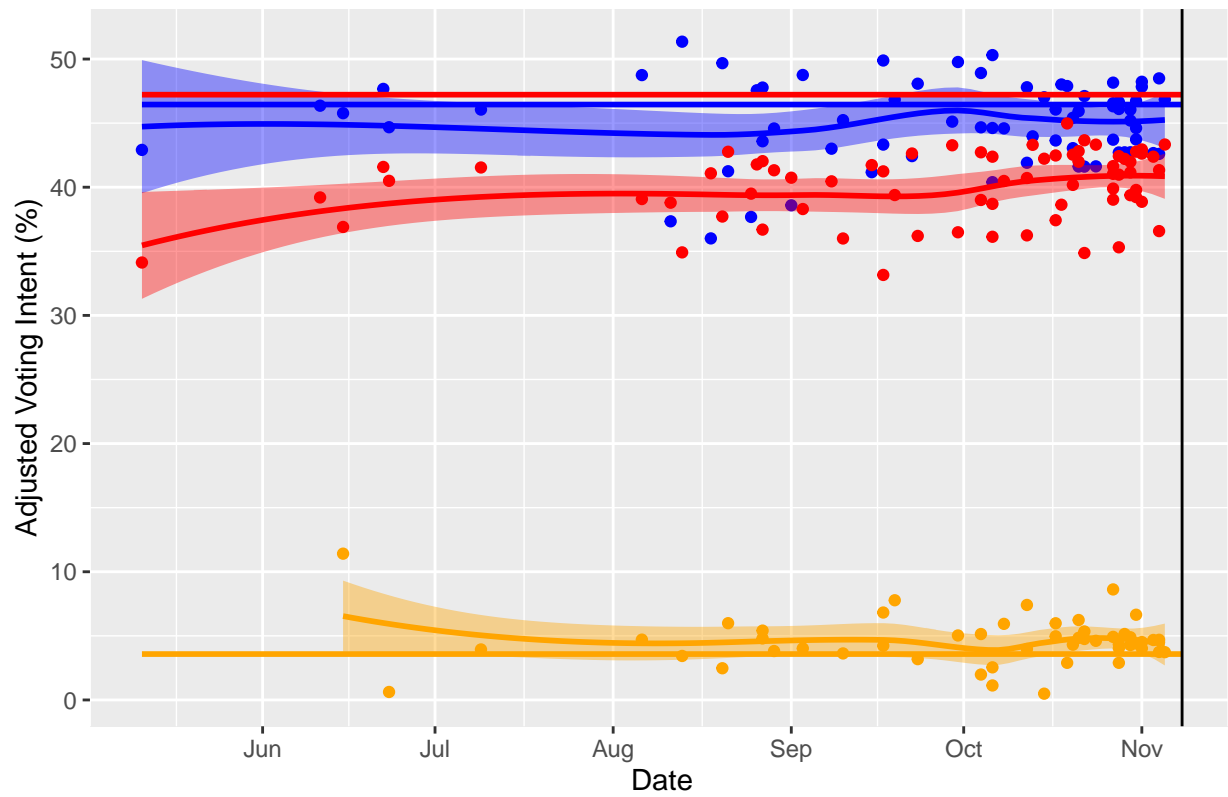




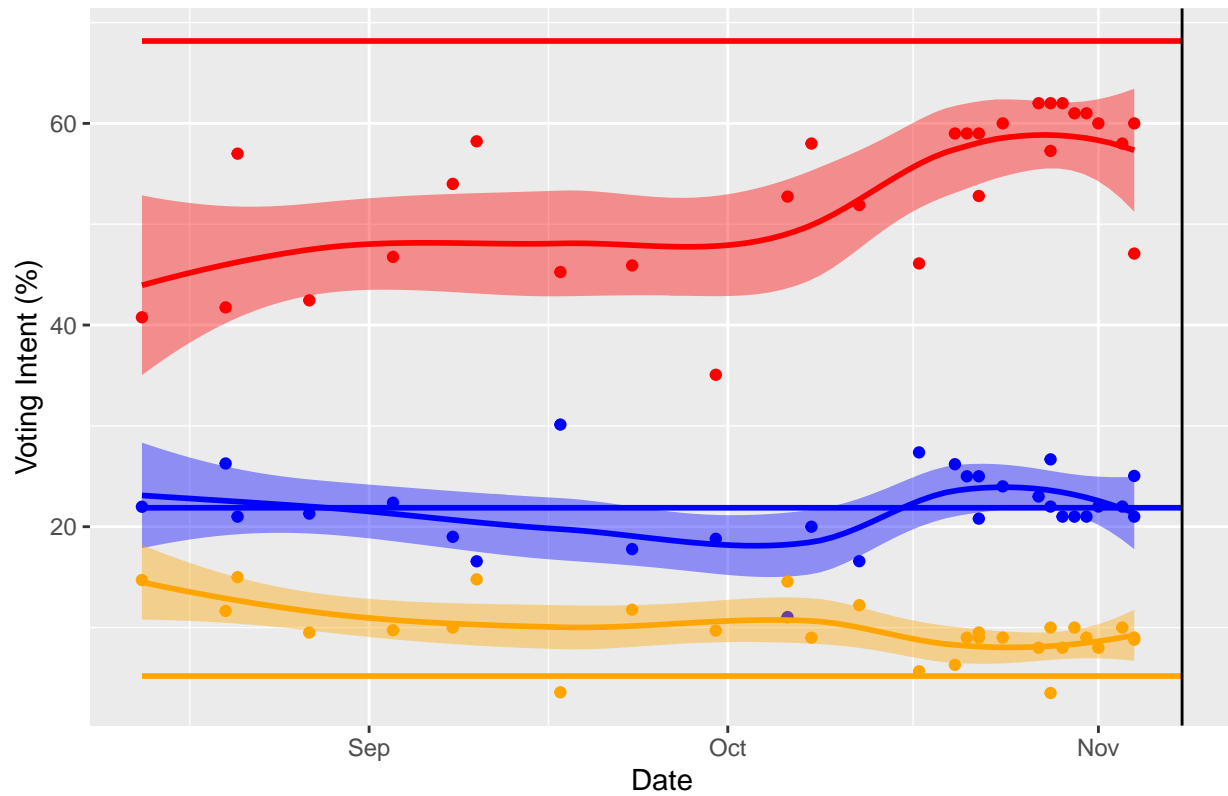
Wisconsin – Raw Polls



Wisconsin – Adjusted Polls



Wyoming – Raw Polls



Wyoming – Adjusted Polls

