```r
knitr::opts_chunk$set(  echo=TRUE, eval=FALSE)
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-26. For overview type 'help("mgcv-package")'.
```

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.0.0      v purrr   0.2.5
## v tibble  1.4.2      v dplyr   0.7.7
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::collapse() masks nlme::collapse()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
```

```r
library(brms);library(ggplot2);library(scales)
```

```
## Loading required package: Rcpp
```

```
## Loading 'brms' package (version 2.7.0). Useful instructions
## can be found by typing help('brms'). A more detailed introduction
## to the package is available through vignette('brms_overview').
## Run theme_set(theme_default()) to use the default bayesplot theme.
```

```
##
## Attaching package: 'brms'
```

```
## The following objects are masked from 'package:mgcv':
##
##     s, t2
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```r
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(directlabels)

##
## Attaching package: 'directlabels'

## The following object is masked from 'package:nlme':
##
##     gapply
```

Estimating Desktop Browser Market Size

Desktop Firefox has been losing users for the past few years and we have also been losing market share. We've generally attributed this to a combination of Chrome taking our users and the generally declining state of the desktop market.

We can use the combination of our user numbers (MAU), which we know, and an external estimate of our market share to estimate the overall size of the market. We have MAU values for each day of the month, but market share only for a month as a whole. As an approximation, we take the MAU value for the last day of the month and attribute it to the month as a whole.

TECHNICAL NOTE: In the current version of this code, we actually assign all these values to the first day of the month rather than the last, which means that all the graphs are actually off by a month or so. This doesn't affect the general trend, but I need to write some code to fix it.

```
ms.ww <- fread("browser-ww-monthly-200901-201910-first.csv")[,date:=as.Date(Date)]
ms.us <- fread("browser-us-monthly-200901-201910-first.csv")[,date:=as.Date(Date)]
mau <- fread("Desktop_MAU_in_US_and_ROW_2019_11_22-last.csv")[,date := as.Date(date)]
adi <- fread("adi-mean.csv")[,date := as.Date(date)]
nm.ww <- fread("netmarketshare.csv")[,date:=as.Date(date)]


# Trim these to the same dates
# First ADI
ms.adi <- ms.ww[date >= min(adi$date) & date <= max(adi$date)]
df.adi <- data.table(date = ms.adi$date, usage = adi$adi, share = ms.adi$Firefox/100)

# Then MAU for StatCounter
ms.ww <- ms.ww[date >= min(mau$date)]
ms.us <- ms.us[date >= min(mau$date)]
mau.st <- mau[date <= max(ms.ww$date)]
df.ww <- data.table(date = ms.ww$date, usage = mau.st$mau_row + mau.st$mau_us, share = ms.ww$Firefox/100
df.us <- data.table(date = ms.us$date, usage = mau.st$mau_us, share = ms.us$Firefox/100)

# Finally MAU for NetMarketShare
mau.nm <- mau[date >= min(nm.ww$date)]
df.nm <- data.table(date = nm.ww$date, usage = mau.nm$mau_row + mau.nm$mau_us, share = nm.ww$Firefox/100
```

Worldwide Market Size

Fhe following figure shows the decline of MAU and market share from the period of 2016-04-01 to 2019-10-01. The MAU data comes from our own telemetry and the market share data comes from StatCounter. Each of these is normalized to the start of the period. Note that market share is declining faster than MAU.

```
df <- df.ww

df$size <- df$usage/df$share
df$usage.normal <- df$usage/df$usage[1]
df$share.normal <- df$share/df$share[1]
df.long <- gather(df, key="Source", value="Frac", usage.normal, share.normal)

res.ww <- df
```

```
ggplot(df.long) + geom_line(mapping=aes(x=date, y=Frac, color=Source))
```
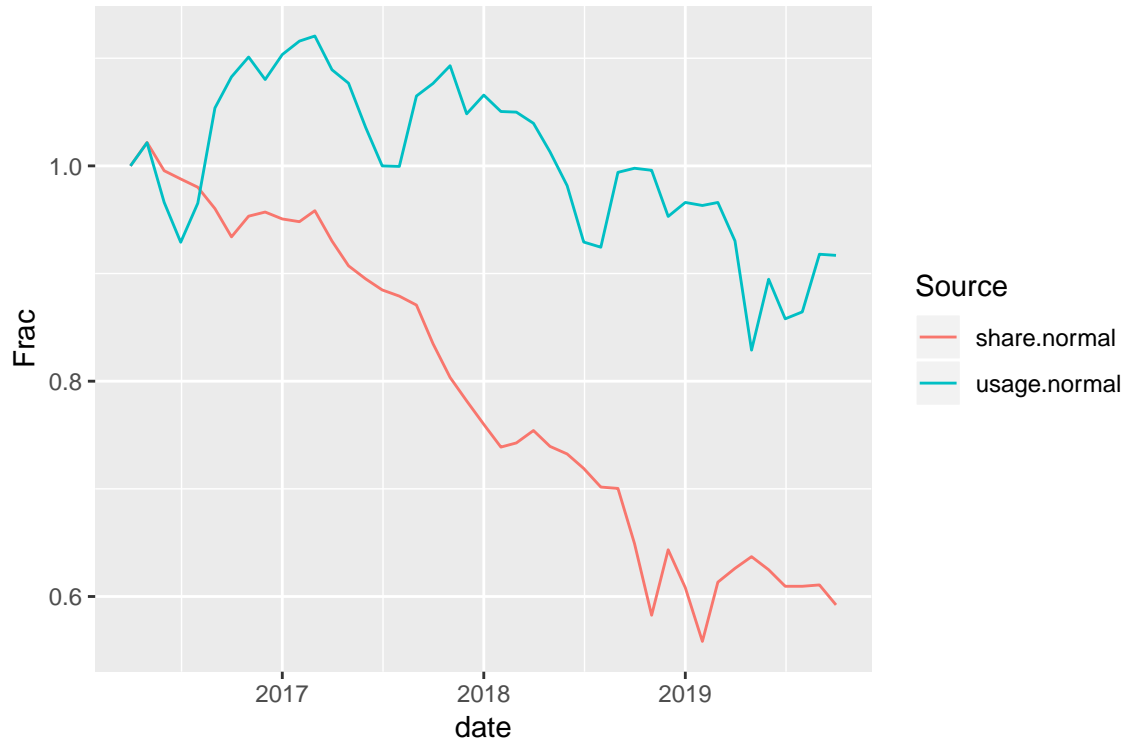

We can get an estimate of the size of the market by dividing MAU by our share of the market, giving us the graph below. Based on this data, desktop market size has actually increased over this period, from about 2 billion to about 3 billion. These numbers do seem fairly high, but it's a straightforward calculation, and we just need to assume that there's no systematic bias in either of these measurements, then it would appear that the market is growing.

```
ggplot(df) + geom_line(mapping=aes(x=date, y=size))
```
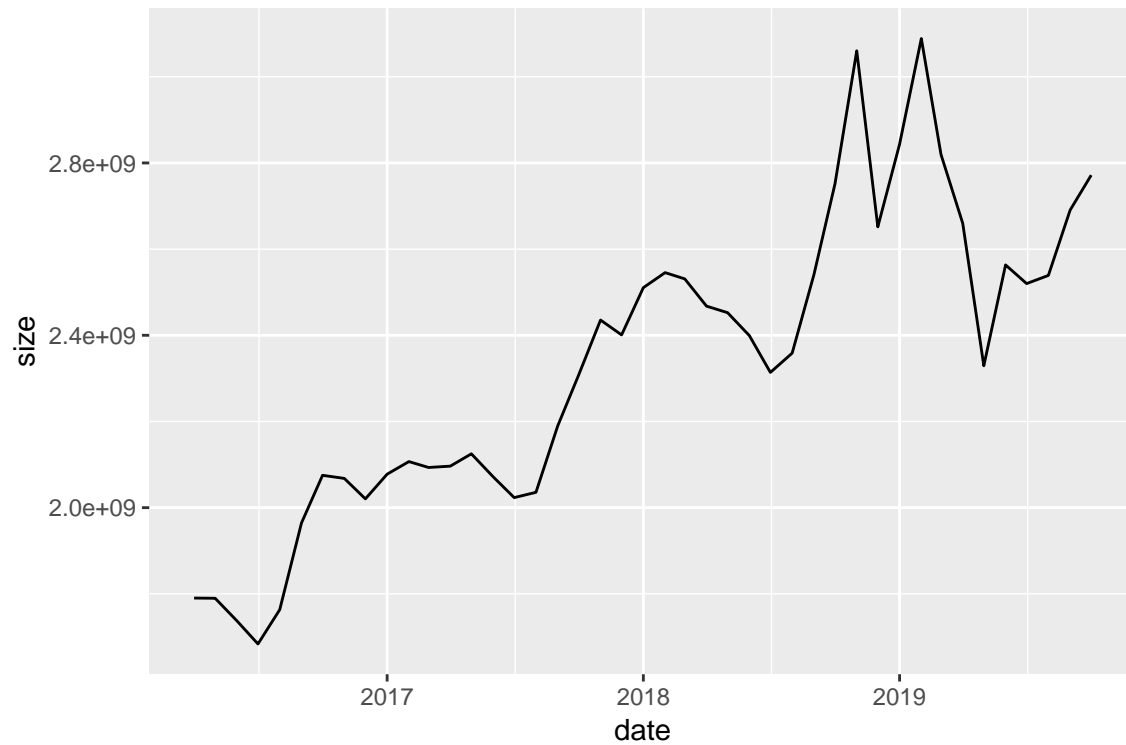
US Market Size

Here is the same treatment for the US, which shows the market size as approximately flat at ~400 million users.

```
df <- df.us

df$size <- df$usage/df$share
df$usage.normal <- df$usage/df$usage[1]
df$share.normal <- df$share/df$share[1]
df.long <- gather(df, key="Source", value="Frac", usage.normal, share.normal)

res.us <-df

ggplot(df.long) + geom_line(mapping=aes(x=date, y=Frac, color=Source))
```
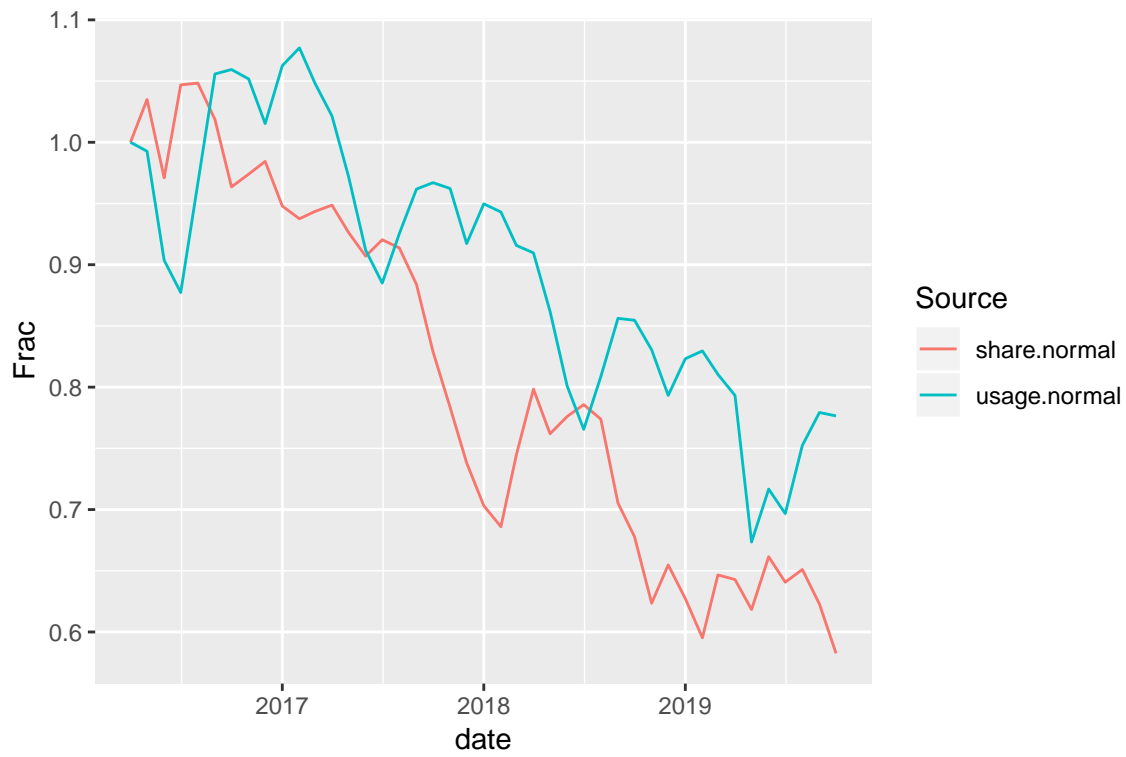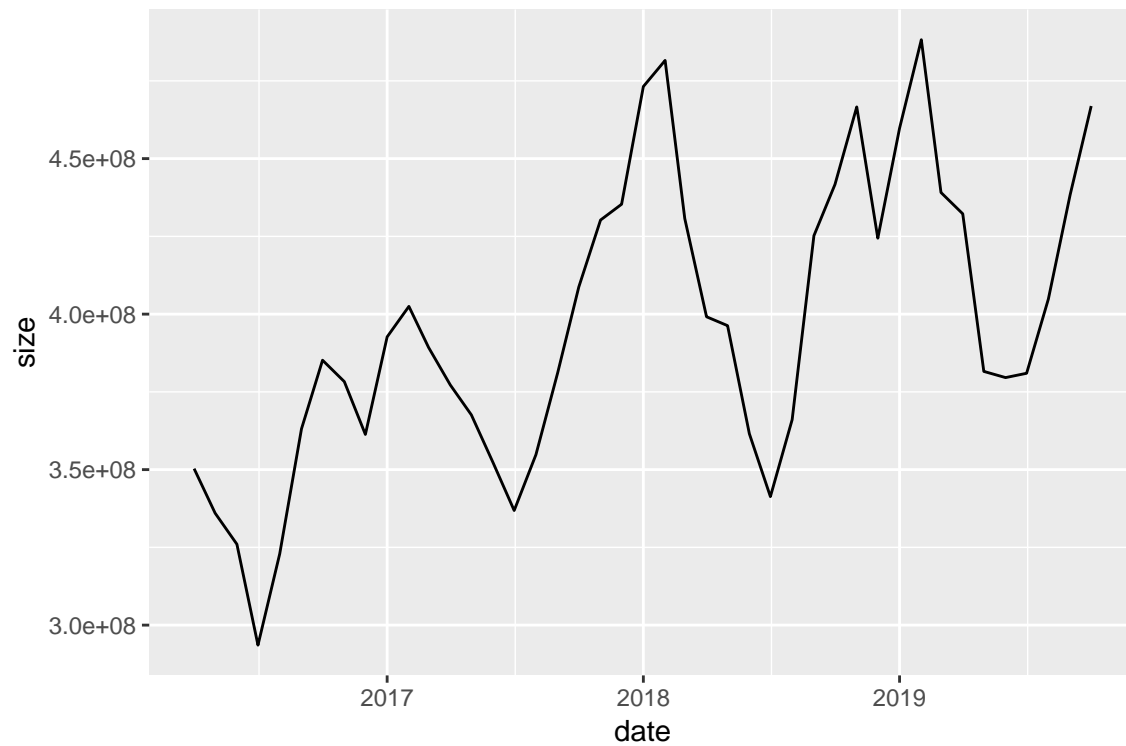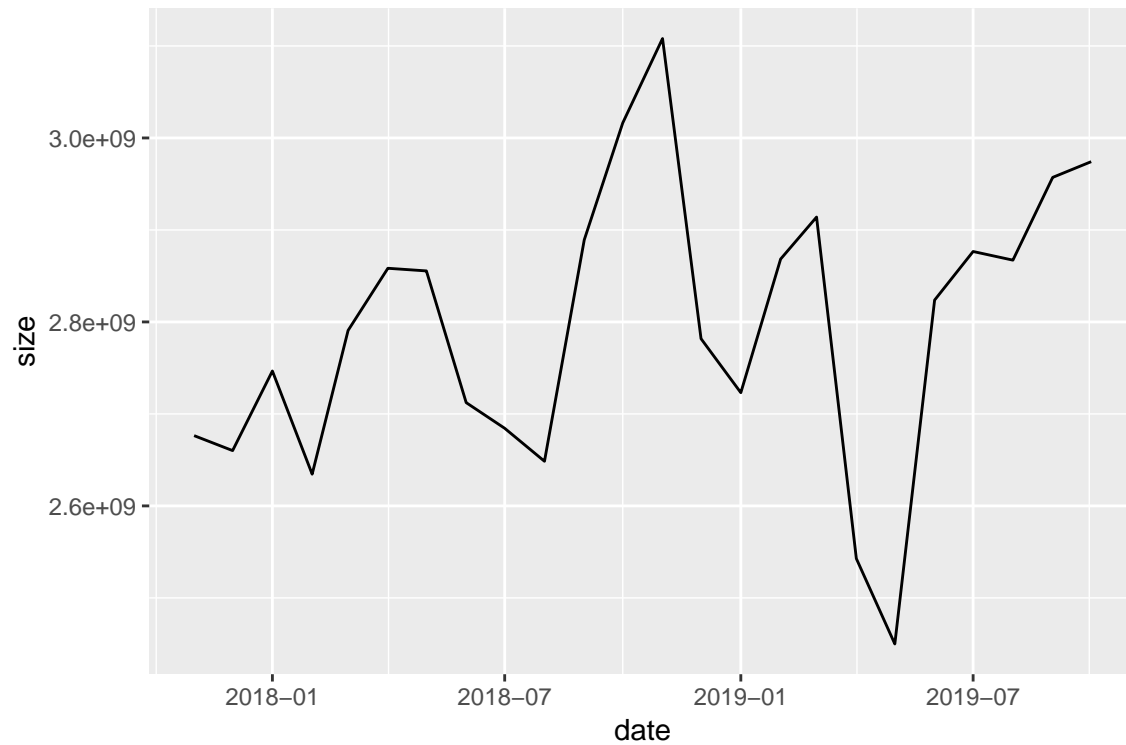
```r
ggplot(df) + geom_line(mapping=aes(x=date, y=size))
```

NetMarketShare

```
df <- df.nm

df$size <- df$usage/df$share
df$usage.normal <- df$usage/df$usage[1]
df$share.normal <- df$share/df$share[1]
df.long <- gather(df, key="Source", value="Frac", usage.normal, share.normal)

res.nm <- df
```

```
ggplot(df.long) + geom_line(mapping=aes(x=date, y=Frac, color=Source))
```



```
ggplot(df) + geom_line(mapping=aes(x=date, y=size))
```

ADI

And finally, here's the same treatment for Average Daily Installs (ADI), which is based on the blocklist ping rather than on telemetry data.
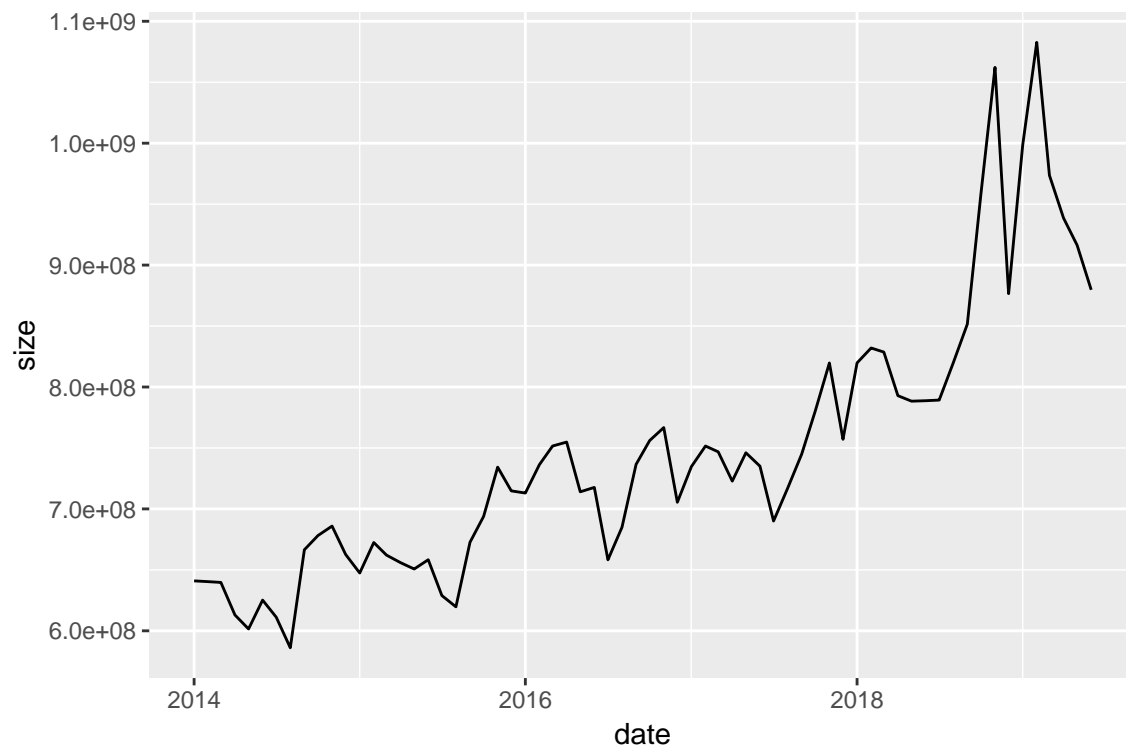
```r
df <- df.adi

df$size <- df$usage/df$share
df$usage.normal <- df$usage/df$usage[1]
df$share.normal <- df$share/df$share[1]
df.long <- gather(df, key="Source", value="Frac", usage.normal, share.normal)

res.adi <- df

ggplot(df.long) + geom_line(mapping=aes(x=date, y=Frac, color=Source))
```
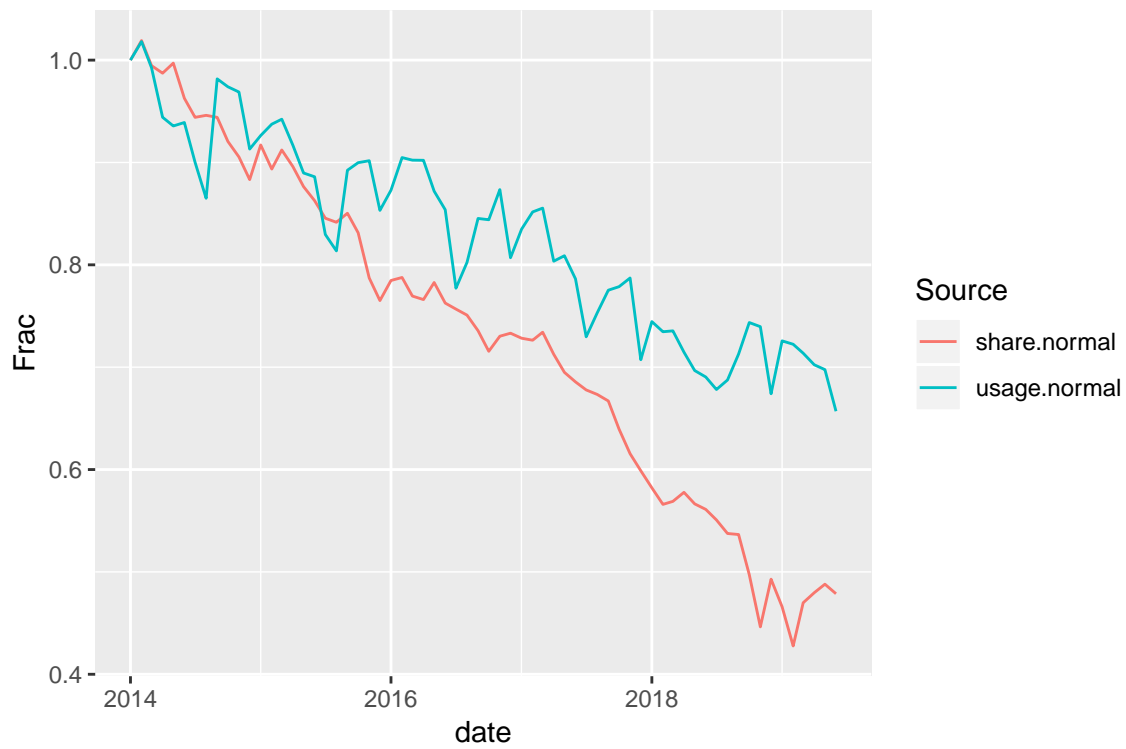
```
ggplot(df) + geom_line(mapping=aes(x=date, y=size))
```

Summary

Finally, we can compare the each of these series by normalizing to the value at the beginning of the MAU series (which starts later than ADI). Because ADI is an unsmoothed value, we take the arithmetic mean of ADI for each month. As before, we attribute it to the first day of the month.

```
res.ww$size.normal <- res.ww$size/res.ww$size[1]
res.us$size.normal <- res.us$size/res.us$size[1]
res.adi$size.normal <- res.adi$size/(res.adi[date == min(res.ww$date)]$size[1])
res.ww$series <- "ww"
res.us$series <- "us"
res.adi$series <- "adi"

res <- rbind(res.ww, res.us, res.adi)
```

```
ggplot(res) + geom_line(mapping=aes(x=date,y=size.normal, color=series)) + geom_smooth(mapping=aes(x=da
```

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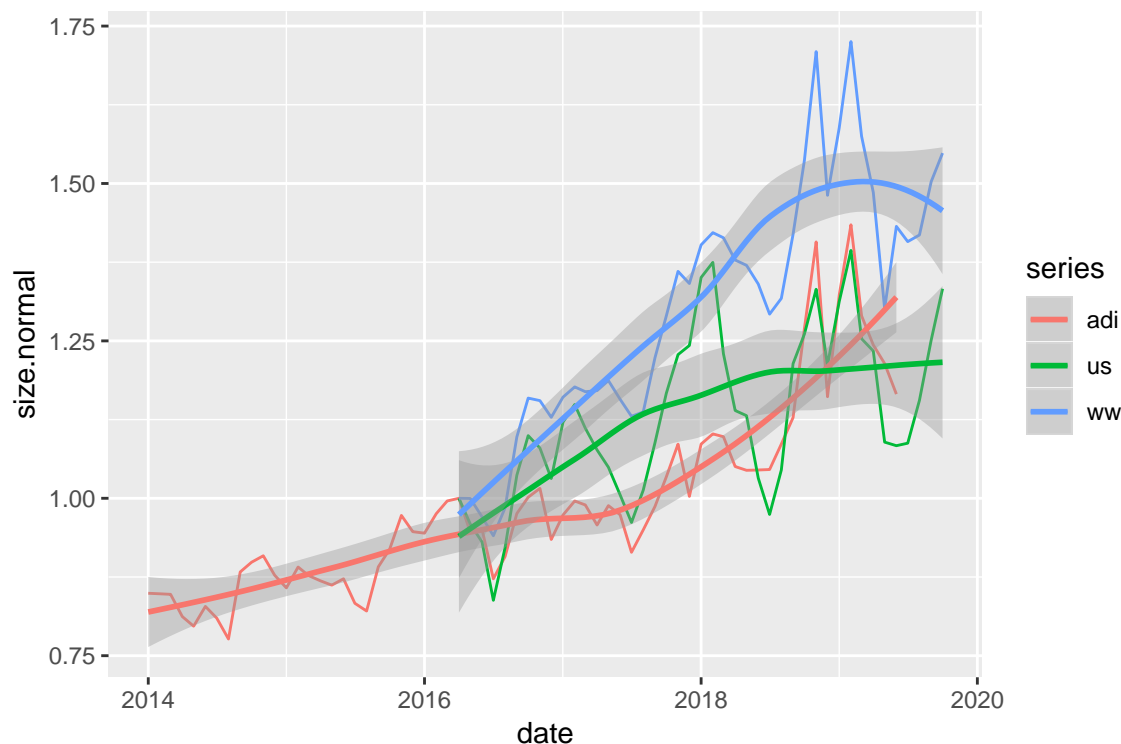

It's probably a mistake to read too much into these data sets (and in particular the projections implied by smoothing are not reliable due to seasonality), but a few points are worth noting. All of this data generally show a market increase over the past few years (causeed by market share decline outpacing absolute decline), coupled with big spikes adound the beginning of 2019. These reflect two big negative spikes in market share around the same time, so may be measurement error in the market share estimates, and in any case will recur for any absolute usage series we run against these market share numbers. We should look to rerun this data with a different source of data, such as NetMarketShare.

In general, this is a somewhat surprising result as we had expected the desktop market to be generally flat to in decline, and so more investigation is definitely required.
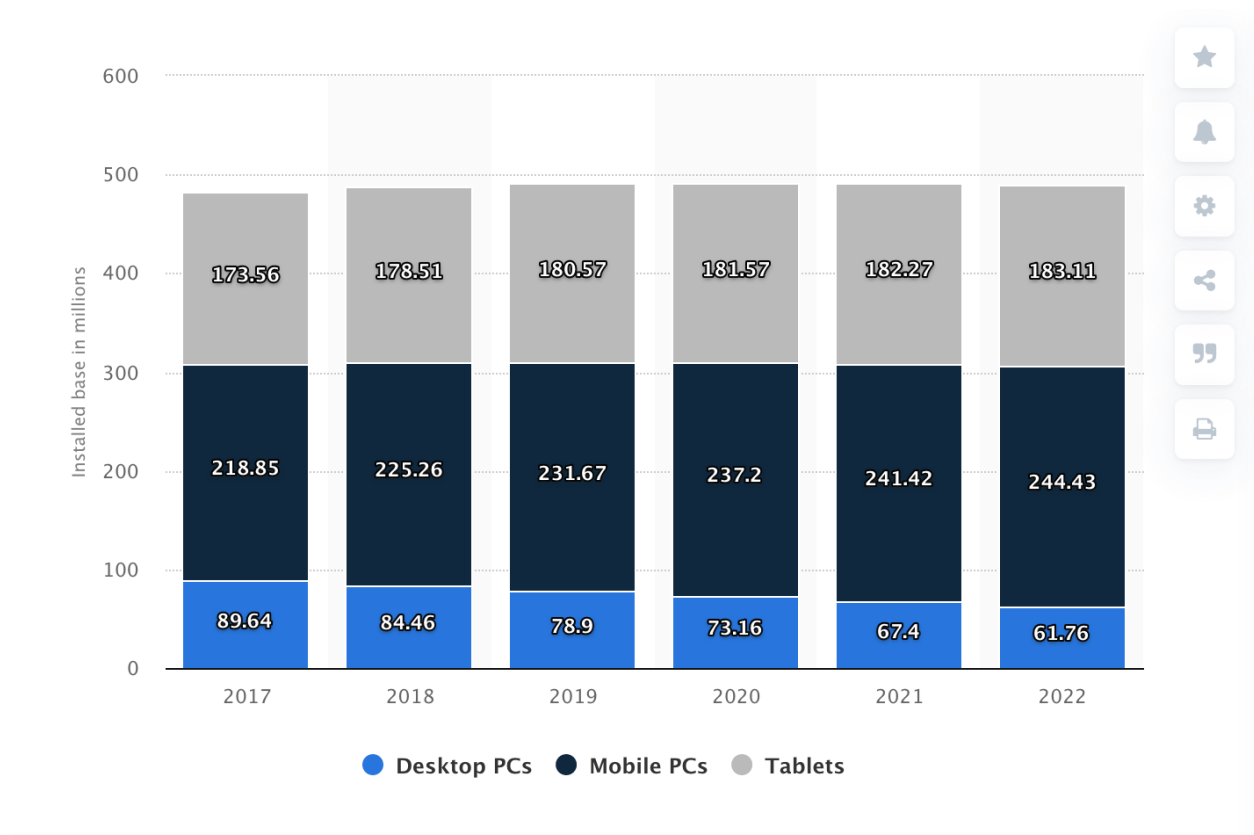
Figure 1: US Desktop and Laptop Market Size (Statista)

Sanity Check

There are a lot of potential sources of noise in this data. As a sanity check, let's compare the US data to an independent source, which is the installed base of desktop and laptops in the US from Statista. [https://www.statista.com/statistics/670172/united-states-installed-base-desktops-laptops-tablets/]. This shows the market as being nearly flat at around 310 million for the past few years, projecting to be 305 million in 2022, so at least we are on the same order of magnitude.