# Datasets

```r
load(file.path(data_dir, "data_ml.RData"))
```

Example Factor, Size

```r
data_ml %>%
   group_by(date) %>% # group by date
   mutate(large = Mkt_Cap_12M_Usd > median(Mkt_Cap_12M_Usd)) %>% # Creates the cap sort
   ungroup() %>% # ungroup
   mutate(year = lubridate::year(date)) %>% # Creates a gear variable
   group_by(year, large) %>% # Analyze by year & cap
   summarize(avg_return = mean(R1M_Usd)) %>% # avg return by year & cap
   ggplot(aes(x = year, y = avg_return, fill = large)) + # plot!
   geom_col(position = "dodge") + # bars side-to-side
   theme(legend.position = c(0.8, 0.2)) + # legend location
   coord_fixed(124) + # x/y aspect ration
   theme(legend.title = element_blank()) +
   scale_fill_manual(values = c("#F87E1F", "#0570EA"), name = "", # colors
                     labels = c("Small", "Large")) +
   ylab("Average Returns") +
   theme(legend.text = element_text(size=9))
```

`summarise()` regrouping output by 'year' (override with `.groups` argument)
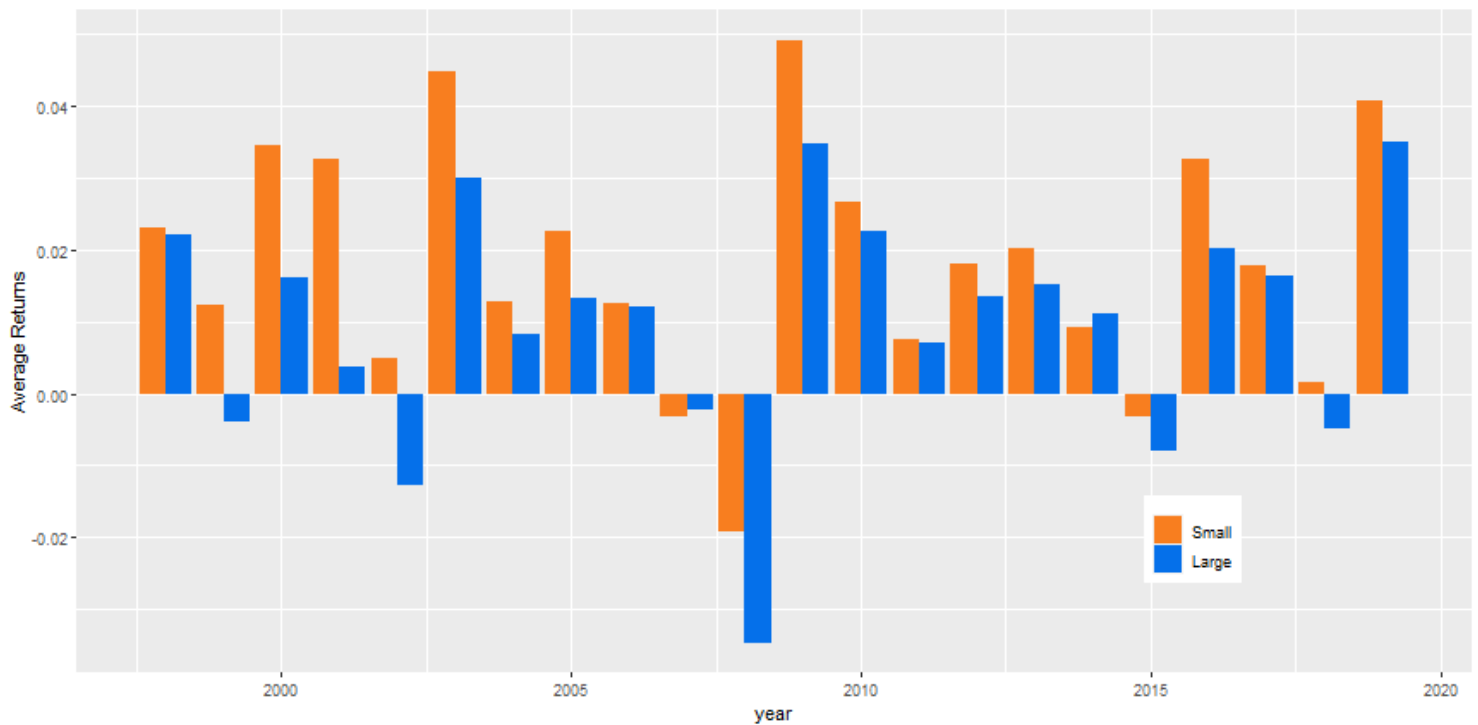
Factors

Size

SMB = small firms minus large firms

Value

HM = high minus low: undervalued minus 'growth' firms

Momentum

WML winners minus losers

Profitability

RMW = robust minus weak profits

profitability is measured as (revenues - (cost and expenses)) / equity

Investment

CMA conservative minus aggressive

Investment is measured via the growth of total assets (divided by total assets).

Low 'risk'

BAB betting against beta

(simple vol, market beta, idiosyncratic vol, etc)

Kenneth French Factor Library

Example Factor Model

```
min_date <- "1963-07-31"; max_date <- "2020-06-30"

temp <- tempfile()

KF_website <- "http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/"
KF_file <- "ftp/F-F_Research_Data_5_Factors_2x3_CSV.zip"

link <- paste0(KF_website, KF_file)

download.file(link, temp, quiet = T)

FF_factors <- read_csv(unz(temp, "F-F_Research_Data_5_Factors_2x3.CSV"),
                    skip = 3) %>%          # Check the number of lines to skip!
    rename(date = X1, MKT_RF = `Mkt-RF`) %>%  # Change the name of first columns
    mutate_at(vars(-date), as.numeric) %>%           # Convert values to number
    mutate(date = ymd(parse_date_time(date, "%Y%m"))) %>%  # Date in right format
    mutate(date = rollback(date + months(1)))        # End of month date
```

```
Warning: Missing column names filled in: 'X1' [1]


-- Column specification -----------------------------------------------------
cols(
  X1 = col_character(),
  `Mkt-RF` = col_character(),
```

```
  SMB = col_character(),
  HML = col_character(),
  RMW = col_character(),
  CMA = col_character(),
  RF = col_character()
)
```

Warning: 1 parsing failure.
row col expected   actual   file
693 -- 7 columns 1 columns <connection>

Warning: Problem with `mutate()` input `MKT_RF`.
i NAs introduced by coercion
i Input `MKT_RF` is `.Primitive("as.double")(MKT_RF)`.

Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion

Warning: Problem with `mutate()` input `SMB`.
i NAs introduced by coercion
i Input `SMB` is `.Primitive("as.double")(SMB)`.

Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion

Warning: Problem with `mutate()` input `HML`.
i NAs introduced by coercion
i Input `HML` is `.Primitive("as.double")(HML)`.

Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion

Warning: Problem with `mutate()` input `RMW`.
i NAs introduced by coercion
i Input `RMW` is `.Primitive("as.double")(RMW)`.

Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion

Warning: Problem with `mutate()` input `CMA`.
i NAs introduced by coercion
i Input `CMA` is `.Primitive("as.double")(CMA)`.

Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion

Warning: Problem with `mutate()` input `RF`.
i NAs introduced by coercion
i Input `RF` is `.Primitive("as.double")(RF)`.

Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion

Warning: Problem with `mutate()` input `date`.
i 58 failed to parse.
i Input `date` is `ymd(parse_date_time(date, "%Y%m"))`.

Warning: 58 failed to parse.

Table 1: Sample of Monthly Factor Returns.

date	MKT_RF	SMB	HML	RMW	CMA	RF
1963-07-31	-0.0039	-0.0045	-0.0094	0.0066	-0.0115	0.0027
1963-08-31	0.0507	-0.0082	0.0182	0.0040	-0.0040	0.0025
1963-09-30	-0.0157	-0.0048	0.0017	-0.0076	0.0024	0.0027
1963-10-31	0.0253	-0.0130	-0.0004	0.0275	-0.0224	0.0029
1963-11-30	-0.0085	-0.0085	0.0170	-0.0045	0.0222	0.0027
1963-12-31	0.0183	-0.0190	-0.0006	0.0007	-0.0030	0.0029

```r
FF_factors <- FF_factors %>% mutate(MKT_RF = MKT_RF / 100,
                                    SMB = SMB / 100,
                                    HML = HML / 100,
                                    RMW = RMW / 100,
                                    CMA = CMA / 100,
                                    RF = RF / 100) %>%
  filter(date >= min_date, date <= max_date)

knitr::kable(head(FF_factors), booktabs = T,
             caption = "Sample of Monthly Factor Returns.")
```

```r
FF_Avg_Returns <- FF_factors %>%
  mutate(date = year(date)) %>%
  gather(key = factor, value = value, - date) %>%
  group_by(date, factor) %>%
  summarise(value = mean(value))
```

```
`summarise()` regrouping output by 'date' (override with `.groups` argument)
```

```r
FF_Avg_Returns %>%
  ggplot(aes(x = date, y = value, color = factor)) +
  geom_line() + coord_fixed(500)
```

```r
FF_factors %>%
  gather(key = factor, value = return, - date) %>%
  filter(factor != 'RF') %>%
  ggplot(aes(return, group = factor)) +
    geom_density(aes(fill = factor, alpha = .25))
```

```r
FF_Avg_Returns %>%
  filter(factor != 'RF') %>%
  ggplot(aes(value, group = factor)) +
    geom_density(aes(fill = factor, alpha = .25))
```
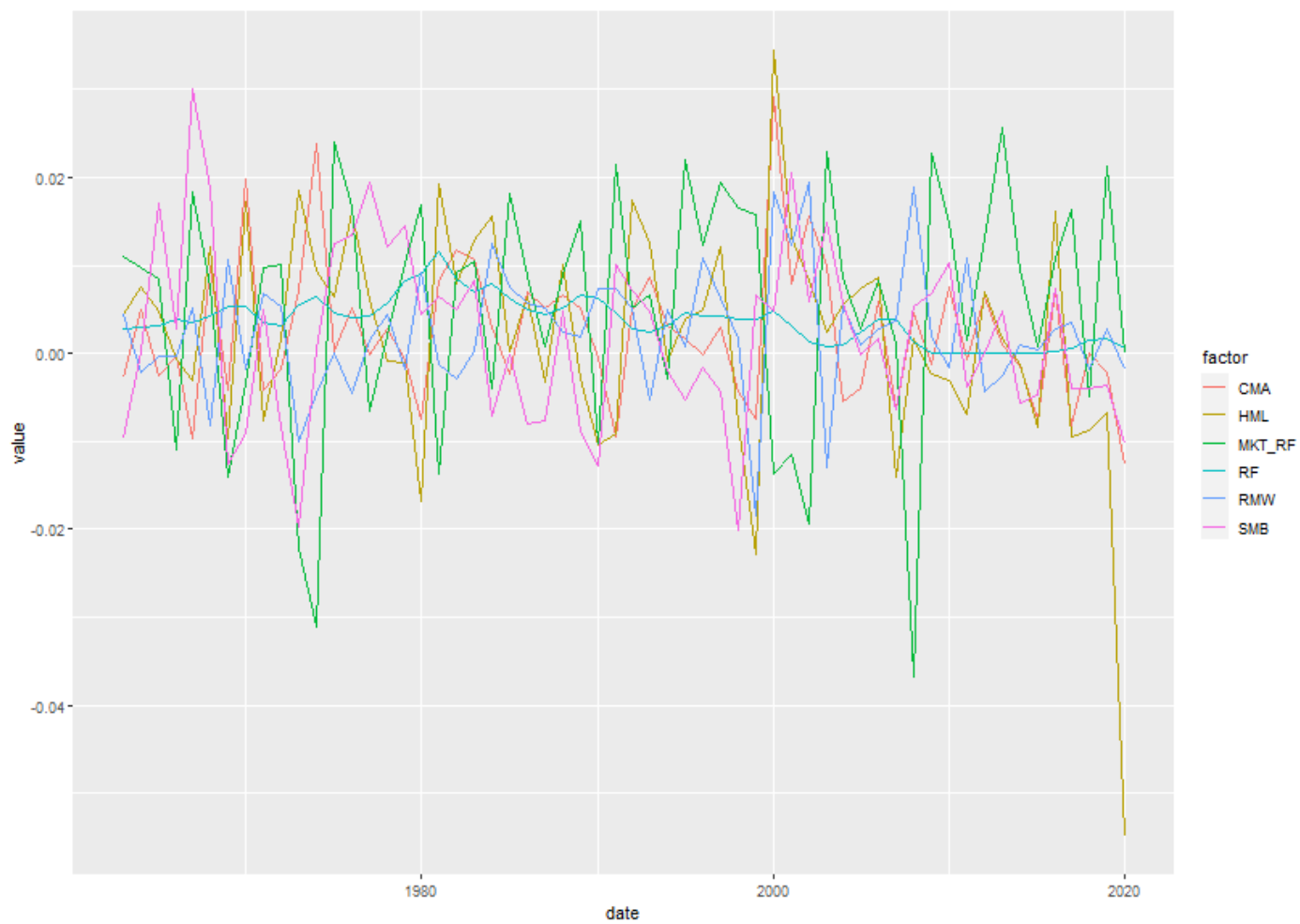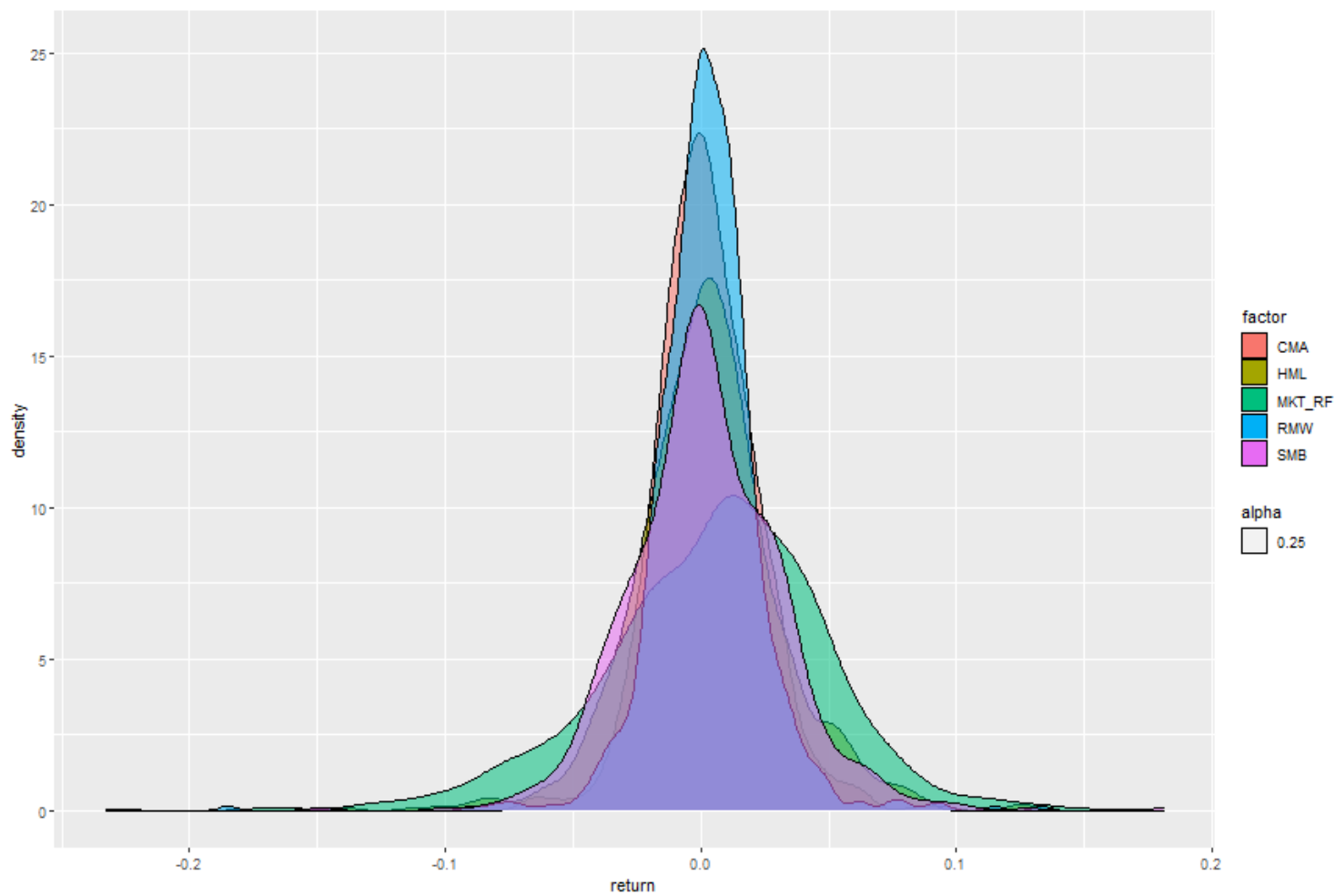
Figure 1: Factor Returns
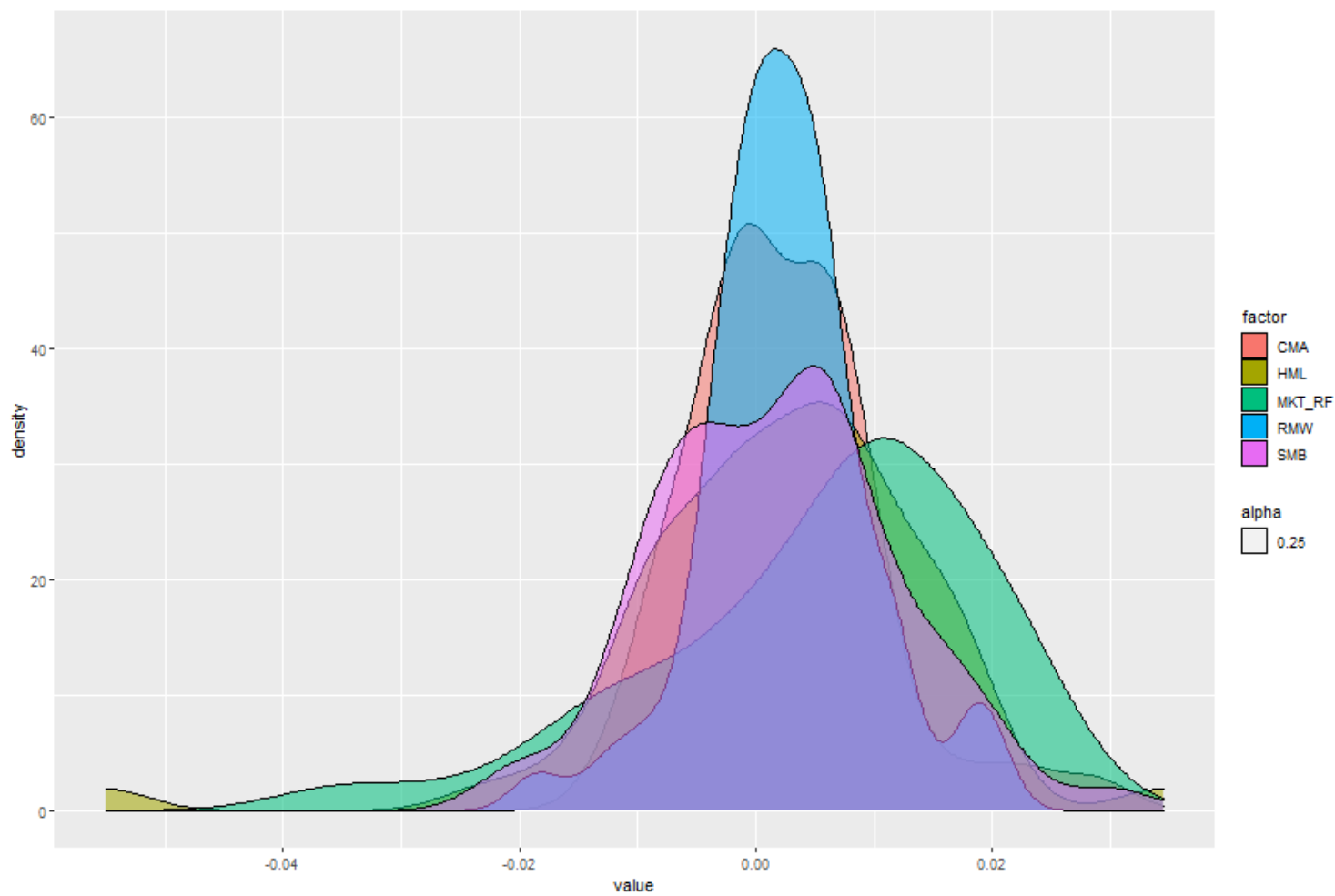
Figure 2: Return Densitites

Figure 3: Yearly Avg Return Densities

```r
FF_Cum_Returns <- FF_factors %>%
    gather(key = factor, value = value, -date) %>%
    group_by(factor) %>%
    mutate(lag_ret = lag(value)) %>%
    mutate(return = cumprod(1 + ifelse(is.na(lag_ret), 0, lag_ret)))

FF_Cum_Returns %>%
    ggplot(aes(date, return, group = factor)) +
    geom_line(aes(col = factor))
```
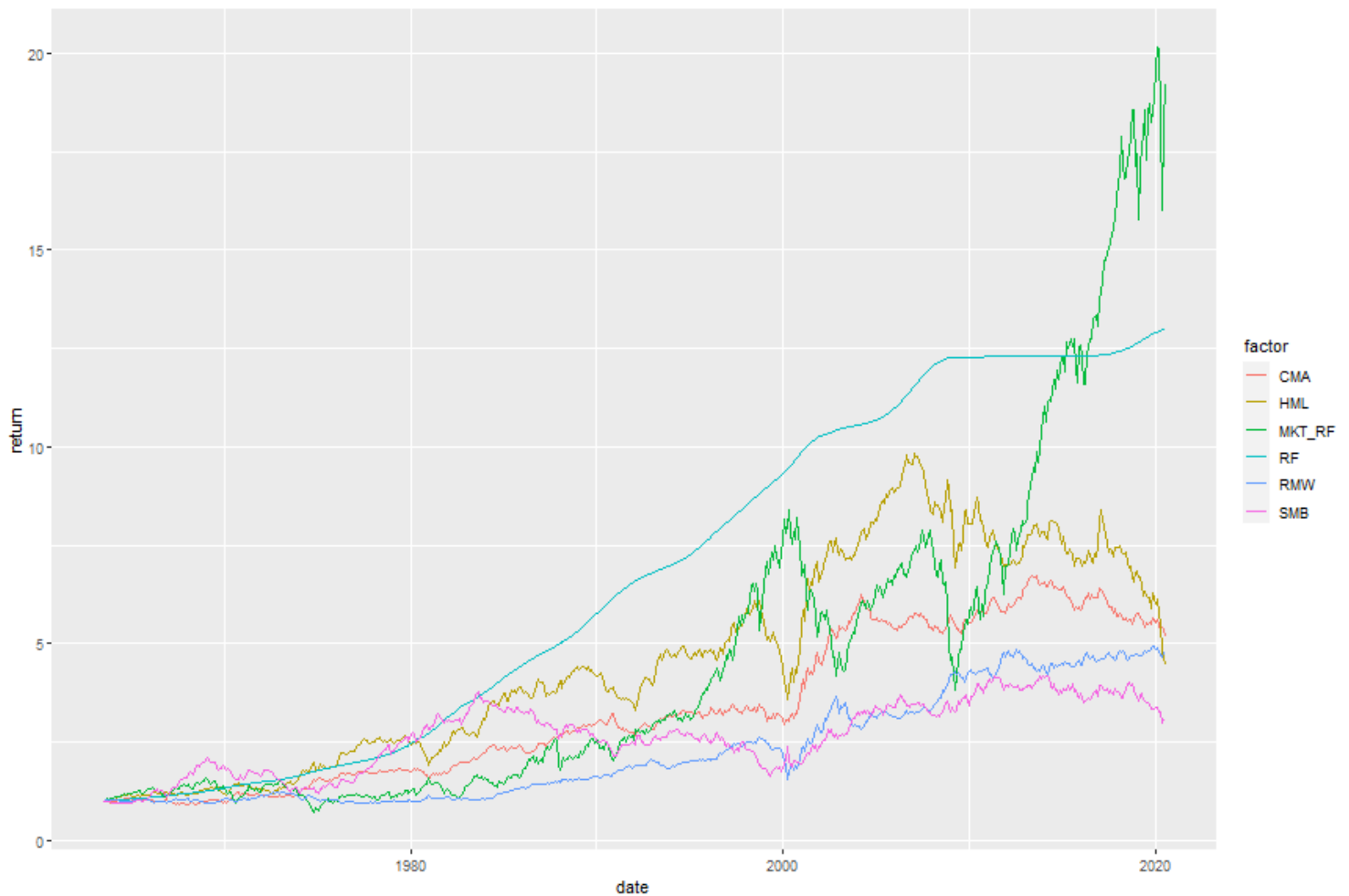

Figure 4: Growth of $1 by factor