

MPF Analysis

Eric Law

18 March, 2021

Prerequisite

```
library(knitr)
library(tidyverse)
library(readxl)
options(warn=-1)
```

Import data

```
MPFA_return = read_excel("data/Fund_Information_Table_risk_return.xlsx", skip = 13)
MPFA_fee = read_excel("data/Fund_Information_Table_fee_charges.xlsx", skip = 13)

# Join two tables
MPFA = inner_join(MPFA_return, MPFA_fee)
```

```
## Joining, by = c("Scheme", "Constituent Fund", "MPF Trustee", "Fund Type", "Latest FER (%)")
```

Data Cleaning

```
# Change the type of data into suitable data type
Return = function(x) {
  ifelse(x == "n.a.", # If the return is "n.a.",
    NA, # save as NA,
    as.numeric(x)) # otherwise, return numeric
}

MPFA = MPFA %>% mutate( # For the return part
  'Launch Date' = 'Launch Date' %>% as.Date(format = "%d-%m-%Y"),
  'Fund size (HKD' m)' = gsub(",", "", 'Fund size (HKD' m)') %>% as.numeric(na.rm = T),
  'Risk Class' = Return('Risk Class'),
  'Latest FER (%)' = Return('Latest FER (%)'),
  'Annualized Return 1 Year (% p.a.)' = Return('Annualized Return 1 Year (% p.a.)'),
  'Annualized Return 5 Year (% p.a.)' = Return('Annualized Return 5 Year (% p.a.)'),
```

```

    'Annualized Return 10 Year (% p.a.)' = Return('Annualized Return 10 Year (% p.a.)'),
    'Annualized Return Since Launch (% p.a.)' = Return('Annualized Return Since Launch (% p.a.)')
) %>% suppressWarnings() # Suppress warnings because of NA

fee = function(x) { # Transform fee from string to number
  x = str_split(x, " ")
  sapply(x, FUN = function(i){# There are 3 cases of fee
    if(length(i) == 1) # Case 1: the fee is exact number
      return(i)
    if(i[2] == "to") # Case 2: up to a number
      return(i[3]) # take the maximum for ease
    else # Case 3: between 2 values
      return(mean(as.numeric(i)[c(1, 3)]) %>% suppressWarnings()) # take the mean for ease
  })
} %>% as.numeric()
}

MPFA = MPFA %>% mutate( #For the fee part
  'Management Fees \n(% p.a.)' = fee('Management Fees \n(% p.a.)'),
  'Administration \nFee/\nTrustee Fee/\nCustodian Fee \n(% p.a.)' = fee('Administration \nFee/\nTrustee Fee/\nCustodian Fee \n(% p.a.)'),
  'Sponsor Fee (% p.a.)' = fee('Sponsor Fee (% p.a.)'),
  'Investment\nManagement \nFee (% p.a.)' = fee('Investment\nManagement \nFee (% p.a.)'),
  'Guarantee Charge \n(% p.a.)' = fee('Guarantee Charge \n(% p.a.)')
)

kable(head(MPFA[, 1:4], 10))

```

Scheme	Constituent Fund	MPF Trustee	Fund Type
AIA MPF - Prime Value Choice	Age 65 Plus Fund	AIAT	Mixed Assets Fund - Default Investment Strategy - Age 65 Plus Fund
AIA MPF - Prime Value Choice	American Fund	AIAT	Equity Fund - United States Equity Fund
AIA MPF - Prime Value Choice	Asian Bond Fund	AIAT	Bond Fund - Asia Bond Fund
AIA MPF - Prime Value Choice	Asian Equity Fund	AIAT	Equity Fund - Asia Equity Fund
AIA MPF - Prime Value Choice	Balanced Portfolio	AIAT	Mixed Assets Fund - 41% to 60% Equity
AIA MPF - Prime Value Choice	Capital Stable Portfolio	AIAT	Mixed Assets Fund - 21% to 40% Equity
AIA MPF - Prime Value Choice	China HK Dynamic Asset Allocation Fund	AIAT	Mixed Assets Fund - Uncategorized Mixed Asset Fund
AIA MPF - Prime Value Choice	Core Accumulation Fund	AIAT	Mixed Assets Fund - Default Investment Strategy - Core Accumulation Fund
AIA MPF - Prime Value Choice	Eurasia Fund	AIAT	Equity Fund - Uncategorized Equity Fund
AIA MPF - Prime Value Choice	European Equity Fund	AIAT	Equity Fund - Europe Equity Fund

Characteristics

Market Shares

```
# Total fund size, grouped by MPF provider
MPFA_market_share = MPFA %>% group_by('MPF Trustee') %>%
  summarise('Constituent Fund' = n(),
            'Total Fund size (HKD' m)' = sum('Fund size (HKD' m)'))

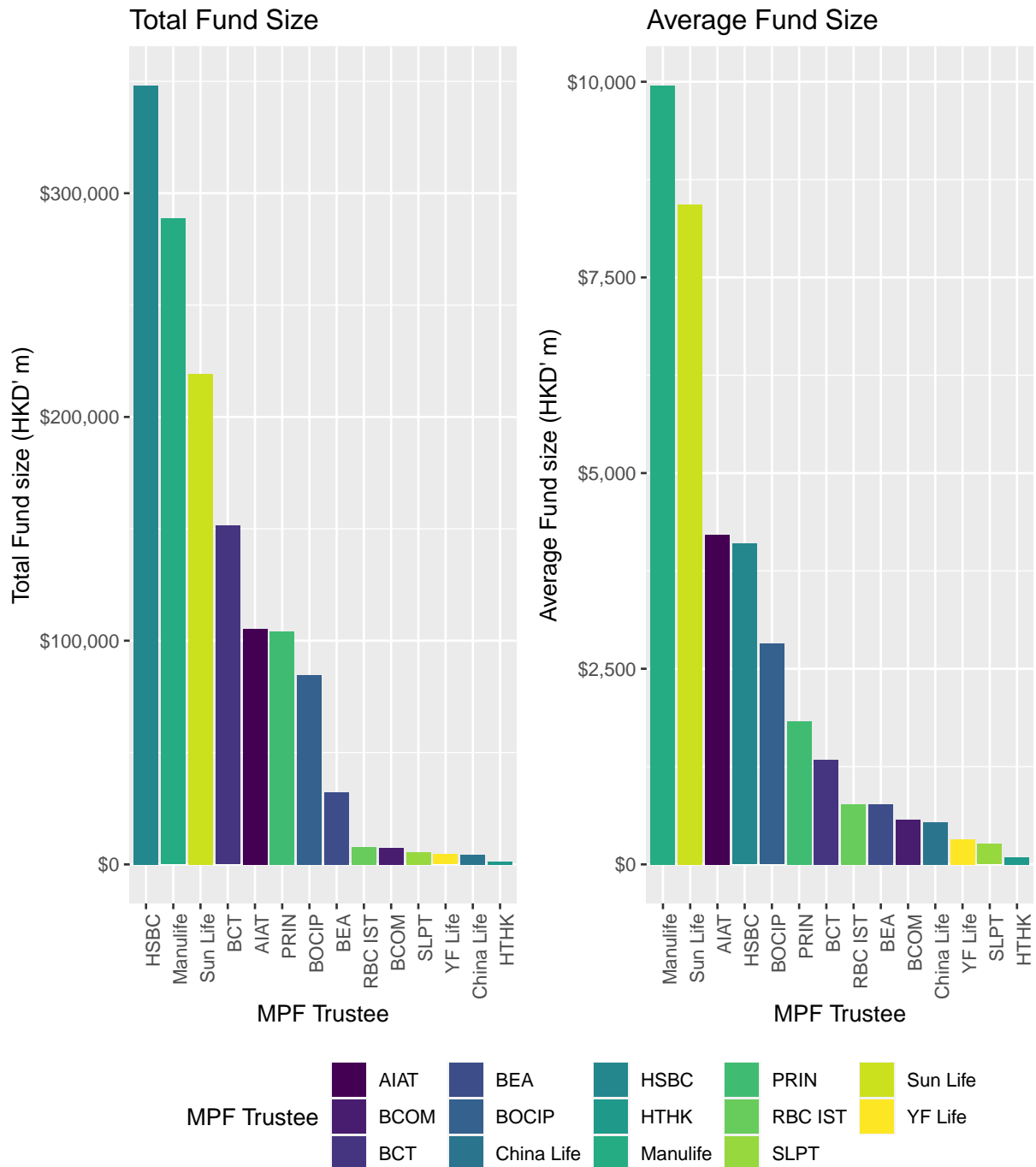
# Adding more columns
MPFA_market_share = MPFA_market_share %>% mutate(
  Proportion = 'Total Fund size (HKD' m)' / sum('Total Fund size (HKD' m)'),
  "Average Fund size (HKD' m)" = 'Total Fund size (HKD' m)' / 'Constituent Fund'
)

# Pie chart to show market share
library(viridis)
library(scales)

bar_total_size = ggplot(MPFA_market_share,
  aes(x = reorder('MPF Trustee', desc('Total Fund size (HKD' m)'), y = 'Total Fund size (HKD' m)',
    fill = 'MPF Trustee')) +
  geom_bar(stat = "identity") +
  scale_fill_viridis(discrete=TRUE) +
  labs(x = "MPF Trustee", title = "Total Fund Size") +
  scale_y_continuous(labels = dollar) +
  theme(axis.text.x = element_text(angle=90, hjust=1))

bar_average_size = ggplot(MPFA_market_share,
  aes(x = reorder('MPF Trustee', desc('Average Fund size (HKD' m)'), y = 'Average Fund size (HKD' m)',
    fill = 'MPF Trustee')) +
  geom_bar(stat = "identity") +
  scale_fill_viridis(discrete=TRUE) +
  labs(x = "MPF Trustee", title = "Average Fund Size") +
  scale_y_continuous(labels = dollar) +
  theme(axis.text.x = element_text(angle=90, hjust=1))
```

```
library(patchwork)
combined = bar_total_size + bar_average_size & theme(legend.position = "bottom")
combined + plot_layout(guides = "collect")
```



We can see some MPF trustees do not have a matched rank of total and average fund size, e.g. HSBC.

Keywords for level of return

```
quadrant = function(x) {
  x = percent_rank(x)
  x = floor(x * 4) / 4
  x = ifelse(x == 1, 0.75, x)
  return(x)
}

MPFA = MPFA %>% mutate(
  "Return 1 Year Quadrant" = quadrant('Annualized Return 1 Year (% p.a.)'),
  "Return 5 Year Quadrant" = quadrant('Annualized Return 5 Year (% p.a.)'),
  "Return 10 Year Quadrant" = quadrant('Annualized Return 10 Year (% p.a.)'),
  "Return Since Launch Quadrant" = quadrant('Annualized Return Since Launch (% p.a.)'),
)

library("tm")
library("SnowballC")
library("wordcloud2")
library("RColorBrewer")

cloud = function(text) {
  docs <- Corpus(VectorSource(text))

  toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
  docs <- tm_map(docs, toSpace, "/")
  docs <- tm_map(docs, toSpace, "@")
  docs <- tm_map(docs, toSpace, "\\|")

  # Convert the text to lower case
  docs <- tm_map(docs, content_transformer(tolower))
  # Remove numbers
  docs <- tm_map(docs, removeNumbers)
  # Remove english common stopwords
  docs <- tm_map(docs, removeWords, stopwords("english"))
  # Remove your own stop word
  # specify your stopwords as a character vector
  docs <- tm_map(docs, removeWords, c("blabla1", "blabla2"))
  # Remove punctuations
  docs <- tm_map(docs, removePunctuation)
  # Eliminate extra white spaces
  docs <- tm_map(docs, stripWhitespace)
  # Text stemming
  # docs <- tm_map(docs, stemDocument)

  dtm <- TermDocumentMatrix(docs)
  m <- as.matrix(dtm)
  v <- sort(rowSums(m), decreasing=TRUE)
  d <- data.frame(word = names(v), freq=v)

  wordcloud2(d)
}
```

```

quad_text = function(Fund, quad) {
  text = list()
  q = seq(0.75, 0, -0.25)
  remove = c("BCOM", "BCT", "Allianz", "AMTD", "MPF", "Fund", "Class", "Sun",
             "Life", "Principal", "Portfolio", "Pro", "Hang Seng", "BEA", "BOC",
             "SHKP", "Schroder", "Haitong", "Fidelity", "Manulife")

  for (i in 1:4) {
    text[[i]] = Fund[quad == q[i] & !is.na(quad)] %>%
      str_remove_all(paste(remove, collapse = "|"))
  }

  return(text)
}

# Check the correlation between risk and returns
library(ggcorrplot)
corr = cor(MPFA[, c(7, 9:11)], use = "complete.obs")
kable(corr)

```

	Risk Class	Annualized Return 1 Year (% p.a.)	Annualized Return 5 Year (% p.a.)	Annualized Return 10 Year (% p.a.)
Risk Class	1.0000000	0.8189506	0.8578981	0.8304491
Annualized Return 1 Year (% p.a.)	0.8189506	1.0000000	0.9212026	0.8242133
Annualized Return 5 Year (% p.a.)	0.8578981	0.9212026	1.0000000	0.9173990
Annualized Return 10 Year (% p.a.)	0.8304491	0.8242133	0.9173990	1.0000000

```
ggcorrplot(corr, type = "lower", lab = T)
```



We can see the risk and returns are highly positively correlated, also highly positively correlated between the returns of different periods.

Since the launching dates of different schemes are not the same, to be fair, let's take 5 years return to look for keywords

```
MPFA_return_q1 = quad_text(MPFA$'Constituent Fund', MPFA$'Return 5 Year Quadrant')
word_cloud = list()
for (i in 1:4)
  word_cloud[[i]] = cloud(MPFA_return_q1[[i]]) %>% suppressWarnings()

library(webshot)
library(htmlwidgets)
```

top 25%:

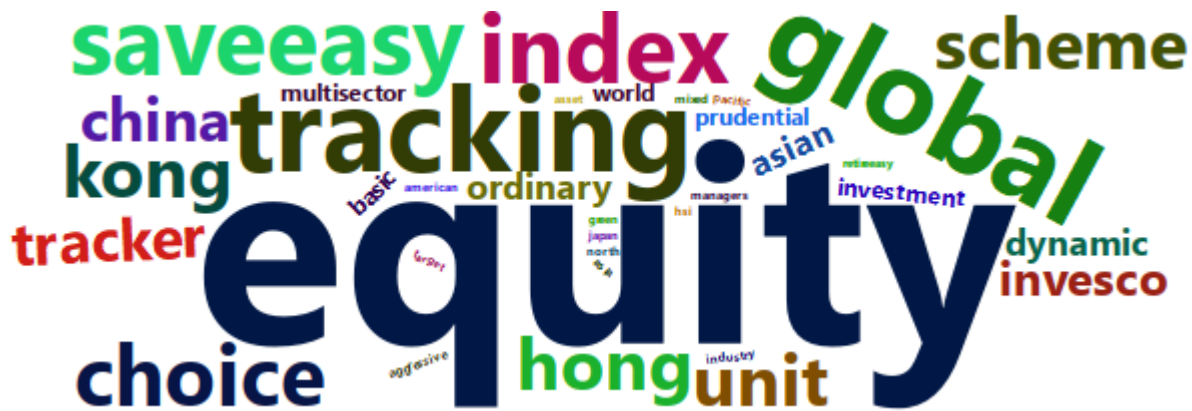
```
min_q1 =

saveWidget(word_cloud[[1]], "wordcloud/q1.html", selfcontained = F)
webshot("wordcloud/q1.html", "wordcloud/q1.png", delay = 5, vwidth = 600, vheight = 200)
```



The keywords of highest return: **China, Greater China, Equity**

q2 - q3:



The keywords of 2nd highest return: **Growth, Balanced, Global, Tracking, Tracking**

A word cloud of financial terms. The word 'stable' is the largest and most prominent, centered at the bottom in a green font. Above it, 'growth' is written in a large orange font. To the right of 'growth', 'capital' is written in a large purple font. Other visible terms include 'equity' (green, top right), 'global' (green, top right), 'asset' (purple, top center), 'scheme' (purple, top center), 'retirement' (purple, top center), 'accumulation' (purple, top left), 'prudential' (blue, left), 'savings' (blue, left), 'flexi' (blue, middle left), 'plus' (blue, middle left), 'japan' (blue, middle left), 'age' (blue, middle left), 'bond' (blue, middle left), 'core' (blue, middle left), 'choice' (blue, middle right), 'unit' (blue, middle right), 'term' (blue, middle right), 'yield' (blue, middle right), 'invesco' (blue, right), 'pacific' (blue, right), 'asia' (blue, bottom center), and 'long' (blue, bottom center). The words are arranged in a circular pattern around the central 'stable' word.

9

Top 10 Return MPF Schemes

First, extract the top 10 (1 year) return MPF schemes

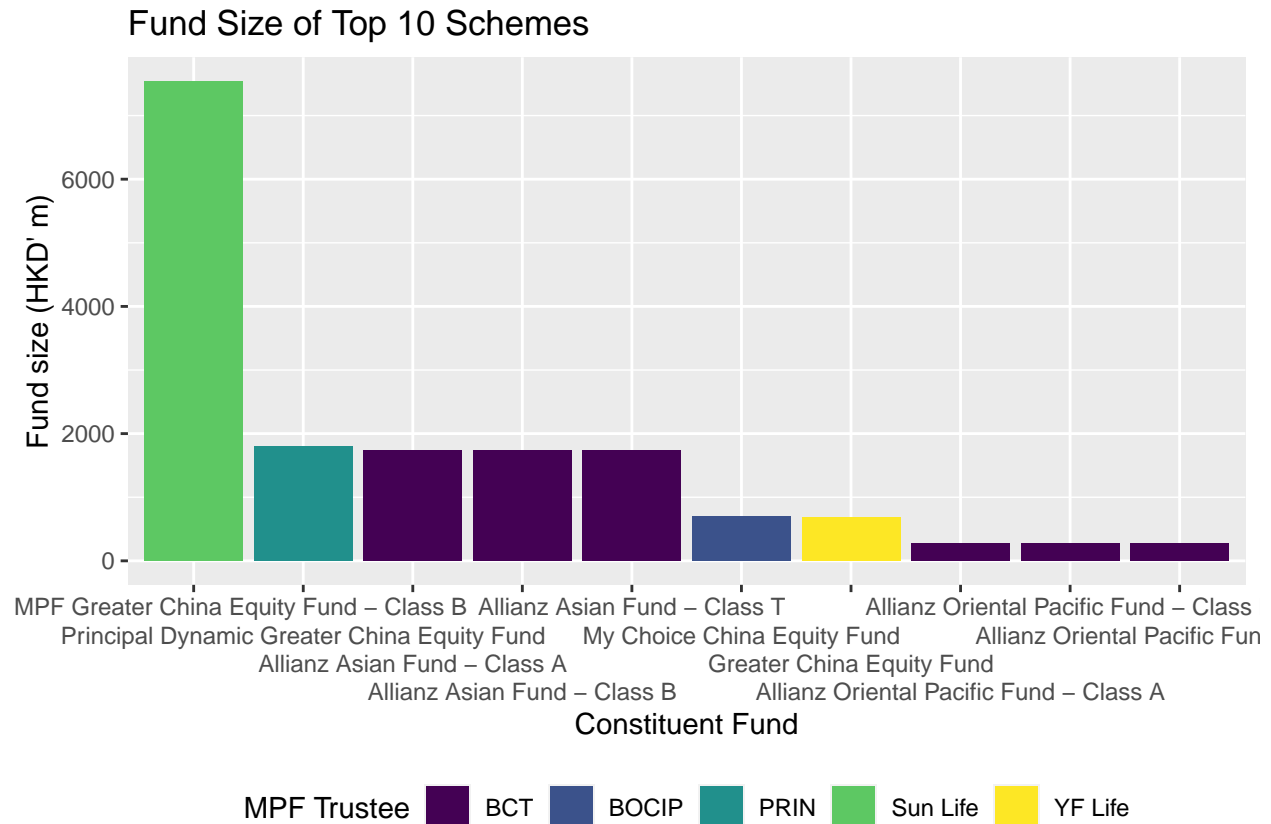
```
MPFA_return_10 = arrange(MPFA, 'Annualized Return 1 Year (% p.a.)' %>% desc()) %>% head(10)
MPFA_return_10 = MPFA_return_10 %>% select(c(2:4, 6:7, 8:11))
kable(MPFA_return_10 %>% head(5))
```

Constituent Fund	MPF Trustee	Fund Type	Fund size (HKD' m)	Risk Class	Latest FER (%)	Annualized Return 1 Year (% p.a.)	Annualized Return 5 Year (% p.a.)	Annualized Return 10 Year (% p.a.)
My Choice China Equity Fund	BOCI	Equity Fund - Greater China Equity Fund	699.37	6	1.09	67.64	24.99	11.68
Allianz Asian Fund - Class T	BCT	Equity Fund - Asia Equity Fund	1730.91	6	1.22	67.12	19.03	9.10
Allianz Asian Fund - Class B	BCT	Equity Fund - Asia Equity Fund	1730.91	6	1.25	67.05	19.09	9.11
Allianz Asian Fund - Class A	BCT	Equity Fund - Asia Equity Fund	1730.91	6	1.45	66.68	18.70	8.86
Greater China Equity Fund	YF Life	Equity Fund - Greater China Equity Fund	689.14	6	1.64	66.53	24.26	NA

We can see the top return schemes are either equity fund or asset fund, also the risk classes are among the highest. Whereas, 6 of which are offered by BCT.

Fund Size of Top 10 Schemes

```
ggplot(MPFA_return_10, aes(reorder('Constituent Fund', 'Fund size (HKD' m)' %>% desc()),
                           'Fund size (HKD' m)', fill = 'MPF Trustee')) +
  geom_bar(stat = "identity") +
  theme(legend.position = "bottom") +
  labs(title = "Fund Size of Top 10 Schemes", x = "Constituent Fund") +
  scale_fill_viridis(discrete=TRUE) +
  scale_x_discrete(guide = guide_axis(n.dodge = 4))
```

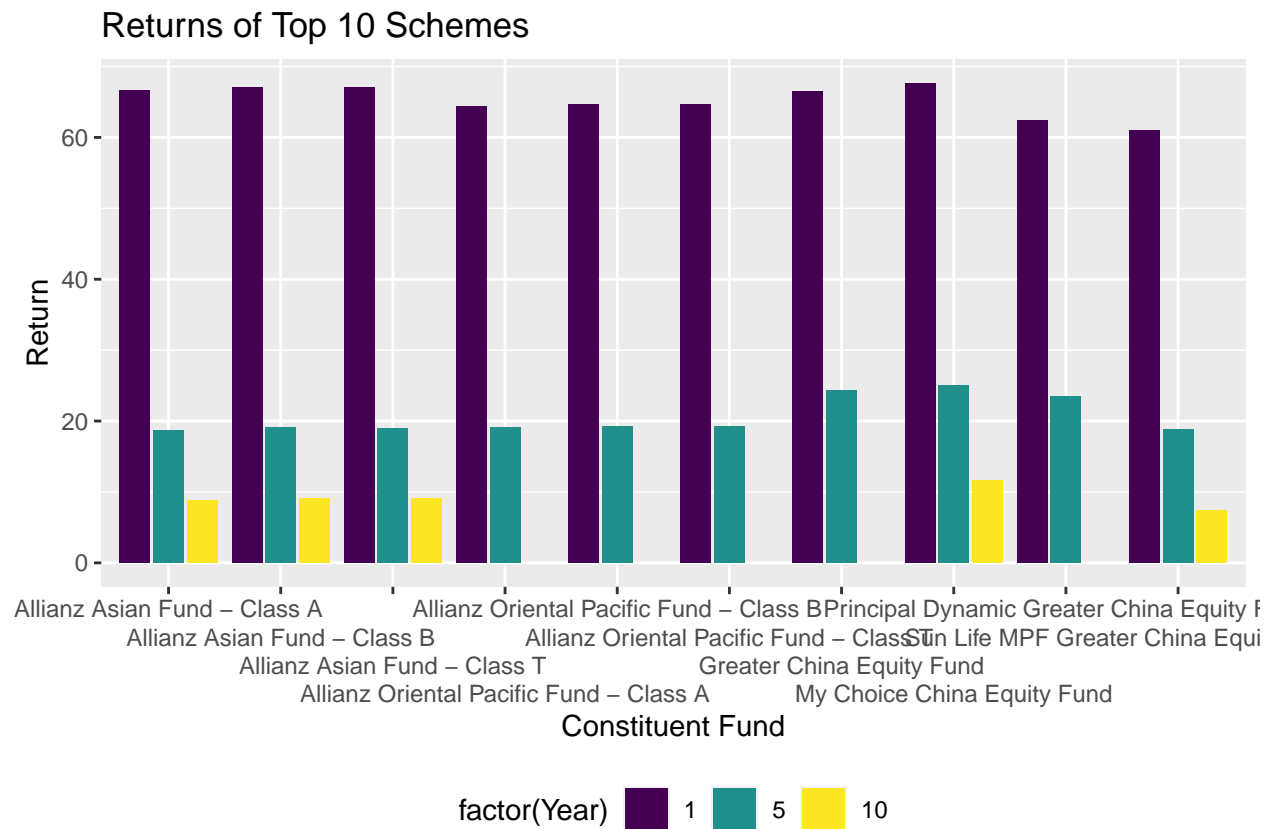


Returns of Top 10 Schemes

```
# Transform the wide data into long data
MPFA_return_10_long = MPFA_return_10[, c(1:2, 7:9)] %>%
  gather("Year", "Return", 3:5, -"MPF Trustee") %>%
  mutate(Year = ifelse(Year == "Annualized Return 1 Year (% p.a.)", 1,
                      ifelse(Year == "Annualized Return 5 Year (% p.a.)", 5, 10))
  )

ggplot(MPFA_return_10_long, aes('Constituent Fund', Return,
                                fill = factor(Year))) +
  geom_bar(stat = "identity", position = "dodge2") +
  scale_x_discrete(guide = guide_axis(n.dodge = 4)) +
```

```
scale_fill_viridis(discrete = TRUE) +
labs(title = "Returns of Top 10 Schemes") +
theme(legend.position = "bottom")
```



Some of the MPF schemes launched less than 10 years, so the 10 years return is missing. The trend of year returns are mostly same.