csr_in_india_data_analysis

Onkar Singh

11/13/2021

https://forcats.tidyverse.org/reference/fct_relevel.html

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

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

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(readxl)
options(scipen = 0)
```

```
state <- readxl::read_excel("CSR_Data_State_Sector_Wise.xlsx" , range = "A2:H40") %>%
  janitor::clean_names()

new_names <- c("state_ut", "FY14-15","FY15-16","FY16-17","FY17-18","FY18-19","FY19-20", "FY20-21")

state <- state %>%
  set_names(new_names) %>%
  pivot_longer(-state_ut, names_to = "financial_year", values_to = "amount_inr_crores") %>%
  mutate(amount_inr_crores = format(amount_inr_crores, scientific = F, digits = 2),
         amount_inr_crores = as.double(amount_inr_crores))
```

```
new_names <- c("sector", "FY14-15","FY15-16","FY16-17","FY17-18","FY18-19","FY19-20", "FY20-21")
sector <- readxl::read_excel("CSR_Data_Development_Sector_Wise.xlsx", range = "A2:H31") %>%
  clean_names() %>%
  set_names(new_names) %>%
  pivot_longer(-sector, names_to = "financial_year", values_to = "amount_inr_crores")
```

Per FY FY14 to FY21

```
sector %>%
  group_by(financial_year) %>%
  summarize(total_csr_spend_per_fy =  sum(amount_inr_crores)) %>%
  mutate(total_csr_spend_per_fy = round(total_csr_spend_per_fy, 0))
```

```
## # A tibble: 7 x 2
##   financial_year total_csr_spend_per_fy
##   <chr>                           <dbl>
## 1 FY14-15                         10066
## 2 FY15-16                         14517
## 3 FY16-17                         14344
## 4 FY17-18                         17098
## 5 FY18-19                         20150
## 6 FY19-20                         24689
## 7 FY20-21                          8828
```

```
state %>%
  group_by(financial_year) %>%
  summarize(total_csr_spend_per_fy =  sum(amount_inr_crores)) %>%
   mutate(total_csr_spend_per_fy = round(total_csr_spend_per_fy, 0))
```

```
## # A tibble: 7 x 2
##   financial_year total_csr_spend_per_fy
##   <chr>                           <dbl>
## 1 FY14-15                         10066
## 2 FY15-16                         14517
## 3 FY16-17                         14344
## 4 FY17-18                         17098
## 5 FY18-19                         20150
## 6 FY19-20                         24689
## 7 FY20-21                          8828
```

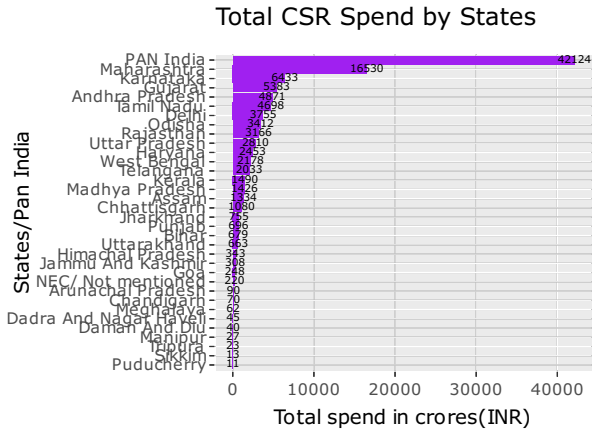
Total FY,Average, Max, Min

```
state %>%
  group_by(financial_year) %>%
  summarize(total_csr_spend_per_fy =  sum(amount_inr_crores)) %>%
  summarize(total_csr_spend = sum(total_csr_spend_per_fy),
            average_csr = mean(total_csr_spend_per_fy),
            max_csr_fy = max(total_csr_spend_per_fy),
            min_csr_fy = min(total_csr_spend_per_fy))
```

```
## # A tibble: 1 x 4
##   total_csr_spend average_csr max_csr_fy min_csr_fy
##             <dbl>       <dbl>      <dbl>      <dbl>
## 1         109692.      15670.     24689.      8828.
```

Per State total CSR Spent from FY14-15 to FY20-21

```
g1 <- state %>%
  filter(amount_inr_crores >= 10) %>%
  group_by(state_ut) %>%
  summarize(total_state_inr_crores = sum(amount_inr_crores)) %>%
  mutate(state_ut = fct_reorder(state_ut, total_state_inr_crores)) %>%
  ggplot(aes(y = state_ut, x = total_state_inr_crores)) +
  geom_col(fill = "purple") +
  geom_text(aes(label = round(total_state_inr_crores,0)), size = 2, hjust = 0, fontface = "bold") +
  labs(y = "States/Pan India",
       x = "Total spend in crores(INR)",
       title = "Total CSR Spend by States", subtitle = "combined total from FY14-15 to FY20-21",
       caption = "data: MCA, India, graph:os2137@caa.columbia.edu")



  plotly::ggplotly(g1)
```

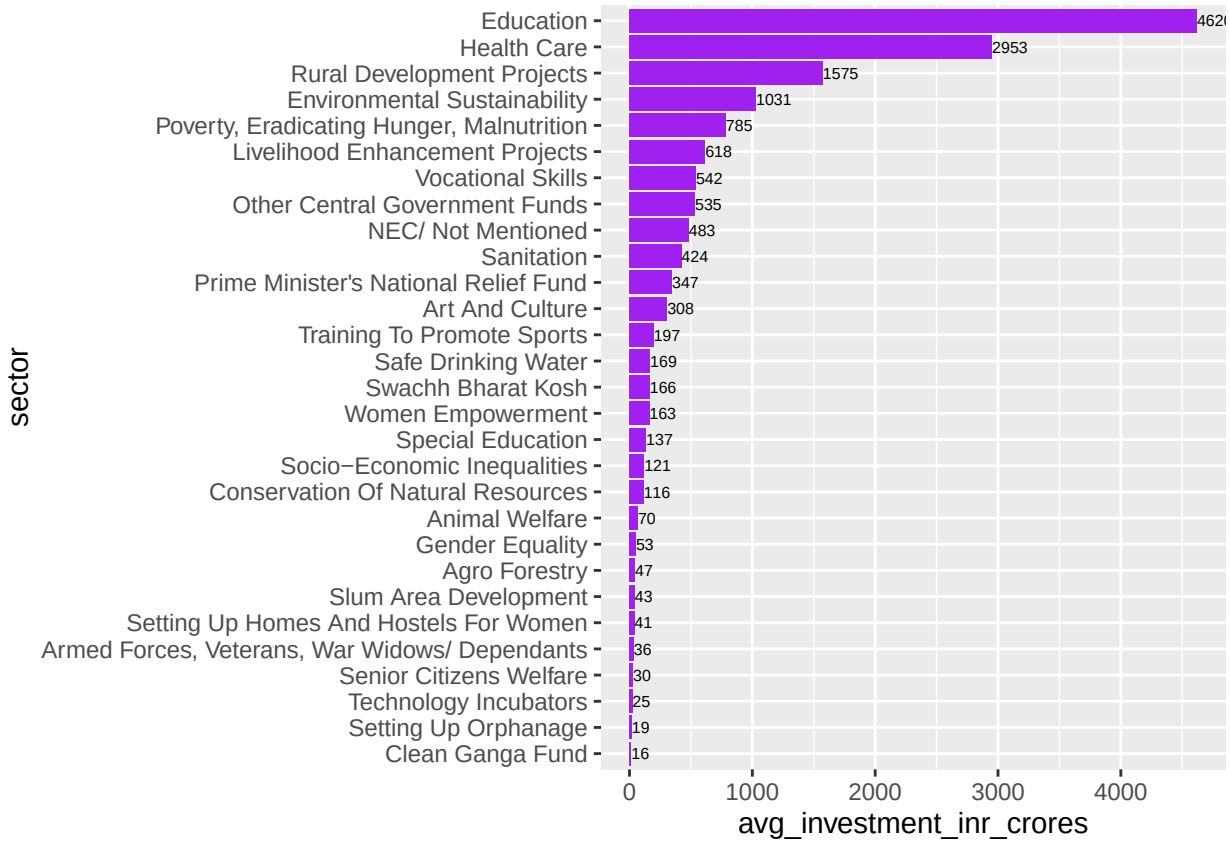


Per State per FY csr spend

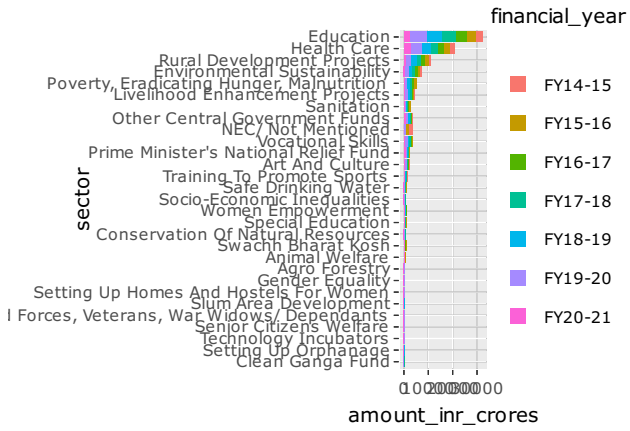
```
g <- state %>%
  # mutate(financial_year = fct_relevel(
  #   financial_year,
  #   c(
  #     "FY14-15",
  #     "FY15-16",
  #     "FY16-17",
  #     "FY17-18",
  #     "FY18-19",
  #     "FY19-20",
  #     "FY20-21"
  #   )
mutate(financial_year = as.factor(financial_year))%>%

  mutate(state_ut = fct_reorder(state_ut, amount_inr_crores)) %>%
  ggplot(aes(y = financial_year, x = amount_inr_crores, fill = forcats::fct_rev(financial_year))) +
  geom_col(position = "dodge") +
geom_text(aes(label = round(amount_inr_crores,0)), size = 2, hjust = 0, vjust = 0.5) +
  facet_wrap(~state_ut)+

scale_fill_discrete(guide=guide_legend(reverse=T))
#   reference: https://stackoverflow.com/questions/38425908/reverse-stacking-order-without-affecting-le
# or +  guides(fill = guide_legend(reverse = TRUE))


# plotly::ggplotly(g)

  # levels(as.factor(state$financial_year))
```

```
g2 <- sector %>%
  group_by(sector) %>%
  summarize(total_investment_inr_crores = sum(amount_inr_crores, na.rm = T)) %>%
  mutate(sector =
           fct_reorder(sector, total_investment_inr_crores)) %>%
  mutate(total_investment_inr_crores = round(total_investment_inr_crores, 0)) %>%
          ggplot(aes(x = total_investment_inr_crores, y = sector)) +
  geom_col(fill = "purple") +
geom_text(aes(label = total_investment_inr_crores), size = 2, hjust = 0)

# plotly::ggplotly(g2)
```

```
sector %>%
  group_by(sector) %>%
  summarize(avg_investment_inr_crores = mean(amount_inr_crores, na.rm = T)) %>%
  mutate(sector =
           fct_reorder(sector, avg_investment_inr_crores)) %>%
  mutate(avg_investment_inr_crores = round(avg_investment_inr_crores, 0)) %>%
          ggplot(aes(x = avg_investment_inr_crores, y = sector)) +
  geom_col(fill = "purple") +

  geom_text(aes(label = avg_investment_inr_crores), size = 2, hjust = 0, vjust = 0.5)
```




```
g <- sector %>%
  mutate(sector = fct_reorder(sector, amount_inr_crores)) %>%
  mutate(amount_inr_crores = round(amount_inr_crores, 0)) %>%
          ggplot(aes(x = amount_inr_crores, y = sector, fill = financial_year)) +
  geom_col() +
  guides(fill = guide_legend(reverse=TRUE))



plotly::ggplotly(g)
```




```
psu_non_psu <- readxl::read_excel("CSR_Data_PSU_Non_PSU_wise.xlsx", range = "A2:O4") %>%
  clean_names()
```

```
names(psu_non_psu)
```

```
##  [1] "psu_non_psu"                     "total_company_fy_2014_15"
##  [3] "amount_spent_fy_2014_15_inr_cr"  "total_company_fy_2015_16"
##  [5] "amount_spent_fy_2015_16_inr_cr"  "total_company_fy_2016_17"
##  [7] "amount_spent_fy_2016_17_inr_cr"  "total_company_fy_2017_18"
##  [9] "amount_spent_fy_2017_18_inr_cr"  "total_company_fy_2018_19"
## [11] "amount_spent_fy_2018_19_inr_cr"  "total_company_fy_2019_20"
## [13] "amount_spent_fy_2019_20_inr_cr"  "total_company_fy_2020_2021"
## [15] "amount_spent_fy_20120_21_inr_cr"
```

```
dim(psu_non_psu)
```

```
## [1]  2 15
```

```
new_names <- c("psu_non_psu",
               "comp_2014to2015",
               "spendinrcrores_2014to2015",
               "comp_2015to2016",
               "spendinrcrores_2015to2016",
               "comp_2016to2017",
```

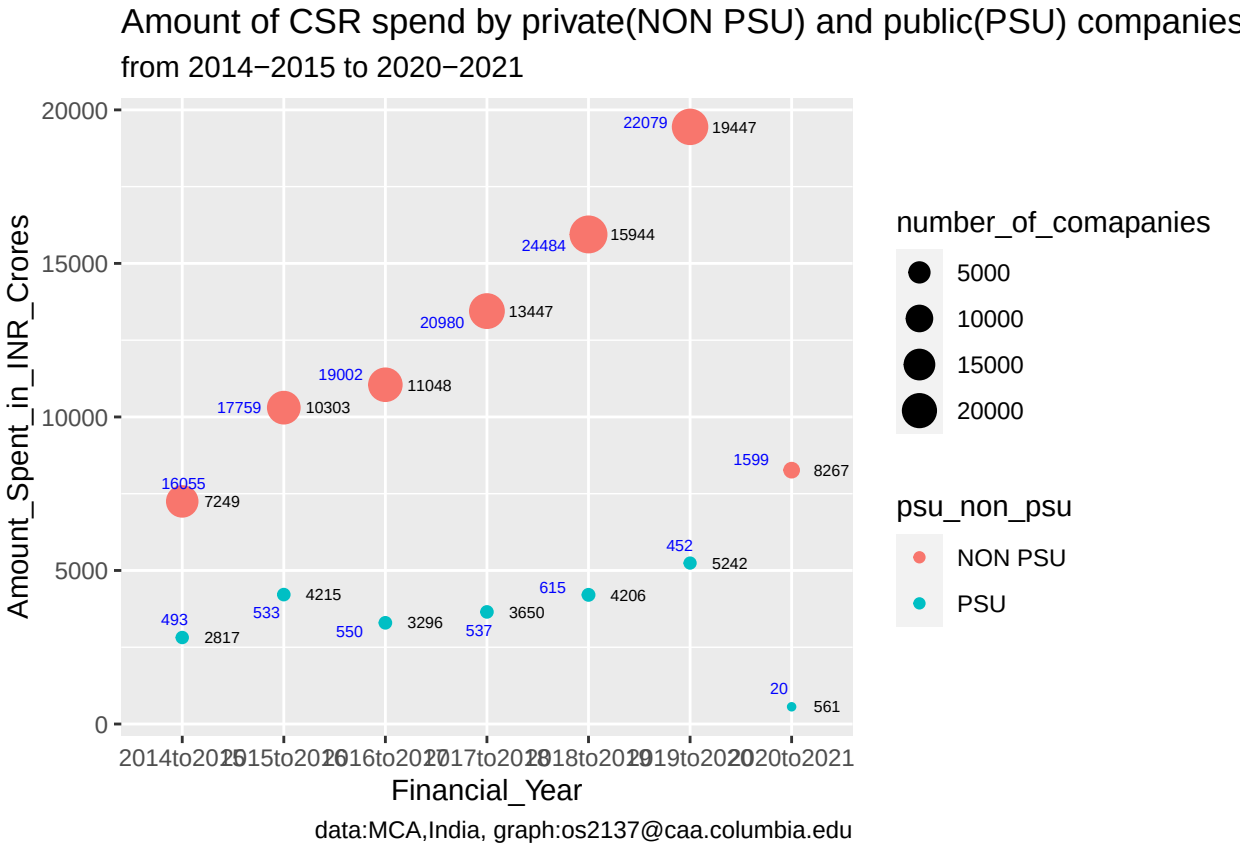
```
                "spendinrcrores_2016to2017",
                "comp_2017to2018",
                "spendinrcrores_2017to2018",
                "comp_2018to2019",
                "spendinrcrores_2018to2019",
                "comp_2019to2020",
                "spendinrcrores_2019to2020",
                "comp_2020to2021",
                "spendinrcrores_2020to2021"
                )
```

```
psu_non_psu <- psu_non_psu %>% set_names(new_names)
psu_non_psu_1 <- psu_non_psu %>%
  select(1, 2,4,6,8,10,12,14) %>%
    pivot_longer(cols = c(comp_2014to2015: comp_2020to2021),
names_to = "company_fy",
values_to = "number_of_comapanies") %>%
  separate(company_fy, c("company", "fy"), sep = "_", extra = "merge")
```

```
psu_non_psu_2 <- psu_non_psu %>%
  select(1, 3, 5, 7, 9, 11, 13, 15) %>%
 pivot_longer( cols = c(spendinrcrores_2014to2015:spendinrcrores_2020to2021),
names_to = "company_fy",
values_to = "amount_spent_inr_crores") %>%
  separate(company_fy, c("company", "fy"), sep = "_", extra = "merge")
```

```
final_psu_non_psu <- psu_non_psu_1 %>%
  left_join(psu_non_psu_2, by = c("psu_non_psu", "fy")) %>%
  select(1, 3, 4, 6)
```

```
final_psu_non_psu %>%
  ggplot(aes(x = fy, y = amount_spent_inr_crores, color = psu_non_psu, size = number_of_comapanies))+
  geom_point() +
  ggrepel::geom_text_repel(aes(label = number_of_comapanies), size = 2, color = "blue", hjust = 0)+
    ggrepel::geom_text_repel(aes(label = round(amount_spent_inr_crores,0)), size = 2, color = "black", h
  labs(x = "Financial_Year",
       y = "Amount_Spent_in_INR_Crores",
       title = "Amount of CSR spend by private(NON PSU) and public(PSU) companies in India",
       subtitle = "from 2014-2015 to 2020-2021",
       caption = "data:MCA,India, graph:os2137@caa.columbia.edu")
```


Amount of CSR spend by private(NON PSU) and public(PSU) companies
from 2014−2015 to 2020−2021
Amount_Spent_in_INR_Crores
20000
15000
10000
5000
0
22079
19447
24484
15944
20980
13447
19002
11048
17759
10303
1599
8267
16055
7249
452
5242
615
4206
533
4215
493
2817
550
3296
537
3650
20
561
2014to2015
2015to2016
2016to2017
2017to2018
2018to2019
2019to2020
2020to2021
Financial_Year
data:MCA,India, graph:os2137@caa.columbia.edu
number_of_comapanies
5000
10000
15000
20000
psu_non_psu
NON PSU
PSU