

Analysis of Space Missions

Ollie Hagger

2023-04-12

R Markdown

Loading Required Libraries

```
library(ggplot2)
library(plotly)
library(gapminder)
library(baseline)
library(plyr)
library(stringr)
library(dplyr)
library(lubridate)
library(data.table)
library(magrittr)
library(forcats)
library(data.table)
```

Loading Data

```
Space_missions <- read.csv("C:/Users/ohagg/OneDrive - University College London/Desktop/Github/Space_mi
```

Check the data structure

```
head(Space_missions)
```

```
##      X Unnamed..0 Company.Name
## 1 0          0      SpaceX
## 2 1          1      CASC
## 3 2          2      SpaceX
## 4 3          3    Roscosmos
## 5 4          4       ULA
## 6 5          5      CASC
##
##                                         Location
## 1 LC-39A, Kennedy Space Center, Florida, USA
## 2 Site 9401 (SLS-2), Jiuquan Satellite Launch Center, China
## 3                               Pad A, Boca Chica, Texas, USA
```

```

## 4           Site 200/39, Baikonur Cosmodrome, Kazakhstan
## 5           SLC-41, Cape Canaveral AFS, Florida, USA
## 6           LC-9, Taiyuan Satellite Launch Center, China
##
##             Datum
## 1 Fri Aug 07, 2020 05:12 UTC
## 2 Thu Aug 06, 2020 04:01 UTC
## 3 Tue Aug 04, 2020 23:57 UTC
## 4 Thu Jul 30, 2020 21:25 UTC
## 5 Thu Jul 30, 2020 11:50 UTC
## 6 Sat Jul 25, 2020 03:13 UTC
##
##                                     Detail Status.Rocket Rocket
## 1 Falcon 9 Block 5 | Starlink V1 L9 & BlackSky StatusActive 50.0
## 2 Long March 2D | Gaofen-9 04 & Q-SAT StatusActive 29.75
## 3 Starship Prototype | 150 Meter Hop StatusActive
## 4 Proton-M/Briz-M | Ekspress-80 & Ekspress-103 StatusActive 65.0
## 5 Atlas V 541 | Perseverance StatusActive 145.0
## 6 Long March 4B | Ziyuan-3 03, Apocalypse-10 & NJU-HKU 1 StatusActive 64.68
##   Status.Mission
## 1   Success
## 2   Success
## 3   Success
## 4   Success
## 5   Success
## 6   Success

```

```
str(Space_missions)
```

```

## 'data.frame': 4324 obs. of  9 variables:
## $ X          : int  0 1 2 3 4 5 6 7 8 9 ...
## $ Unnamed..0  : int  0 1 2 3 4 5 6 7 8 9 ...
## $ Company.Name: chr  "SpaceX" "CASC" "SpaceX" "Roscosmos" ...
## $ Location    : chr  "LC-39A, Kennedy Space Center, Florida, USA" "Site 9401 (SLS-2), Jiuquan Sate
## $ Datum       : chr  "Fri Aug 07, 2020 05:12 UTC" "Thu Aug 06, 2020 04:01 UTC" "Tue Aug 04, 2020 23:57 UTC" "Thu Jul 30, 2020 21:25 UTC" "Thu Jul 30, 2020 11:50 UTC" "Sat Jul 25, 2020 03:13 UTC"
## $ Detail      : chr  "Falcon 9 Block 5 | Starlink V1 L9 & BlackSky" "Long March 2D | Gaofen-9 04 & Q-SAT" "Starship Prototype | 150 Meter Hop" "Proton-M/Briz-M | Ekspress-80 & Ekspress-103" "Atlas V 541 | Perseverance"
## $ Status.Rocket: chr  "StatusActive" "StatusActive" "StatusActive" "StatusActive" "StatusActive" ...
## $ Rocket       : chr  "50.0" "29.75" "65.0" ...
## $ Status.Mission: chr  "Success" "Success" "Success" "Success" ...

```

```
View(Space_missions)
```

Remove the a column from the data table

```
Space_missions <- Space_missions %>% select(-c(X))
```

Date formatting

```
Space_missions <- Space_missions %>%
  mutate(launch_date = as_date(parse_date_time(Datum, c("mdy HM", "mdy"), tz = "UTC")))
```

```
##Rename column names

Space_missions <- Space_missions %>%
  rename(row_names=Unnamed..0)
```

Data Summary and NAs check

```
summary(Space_missions)

##      row.names    Company.Name       Location        Datum
##  Min.   : 0    Length:4324    Length:4324    Length:4324
##  1st Qu.:1081  Class  :character  Class  :character  Class  :character
##  Median :2162   Mode   :character  Mode   :character  Mode   :character
##  Mean   :2162
##  3rd Qu.:3242
##  Max.   :4323

##      Detail      Status.Rocket      Rocket      Status.Mission
##  Length:4324    Length:4324    Length:4324    Length:4324
##  Class  :character  Class  :character  Class  :character  Class  :character
##  Mode   :character  Mode   :character  Mode   :character  Mode   :character
##
##      launch_date
##  Min.   :1957-10-04
##  1st Qu.:1972-04-19
##  Median :1984-12-16
##  Mean   :1987-11-28
##  3rd Qu.:2002-09-10
##  Max.   :2020-08-07

sapply(Space_missions, function(x) sum(is.na(x)))
```

	row.names	Company.Name	Location	Datum	Detail
##	0	0	0	0	0
##	Status.Rocket	Rocket	Status.Mission	launch_date	
##	0	0	0	0	

Extracting country from location

```
Space_missions <- Space_missions %>%
  mutate(country = word(Location,-1))
```

```
Space_missions %>% count(country, sort =T)
```

```
##      country     n
##  1      Russia 1395
```

```

## 2          USA 1344
## 3 Kazakhstan  701
## 4      France  303
## 5      China  268
## 6      Japan  126
## 7      India   76
## 8     Ocean   36
## 9      Iran   13
## 10 Zealand   13
## 11 Israel    11
## 12 Kenya     9
## 13 Korea     8
## 14 Australia  6
## 15 Mexico    4
## 16 Sea       4
## 17 Brazil    3
## 18 Canaria   2
## 19 Facility   1
## 20 Site      1

```

Excluding Non-Country Locations

```

loc <- Space_missions %>%
  select(country, Location) %>%
  filter(country %in% c("Ocean", "Sea", "Facility", "Site"))

View(loc)

```

Renaming Country Names and Removing Extra Information

```

Space_missions <-
  Space_missions %>% mutate(
    country = case_when(
      Location == "LP Odyssey, Kiritimati Launch Area, Pacific Ocean" ~ "Pacific Ocean",
      Location == "LP-41, Kauai, Pacific Missile Range Facility" ~ "Range Facility",
      Location == "K-84 Submarine, Barents Sea Launch Area, Barents Sea" |
        # OR
      Location == "K-496 Submarine, Barents Sea Launch Area, Barents Sea" |
        # OR
      Location == "K-407 Submarine, Barents Sea Launch Area, Barents Sea" ~ "Barents Sea",
      Location == "Tai Rui Barge, Yellow Sea" ~ "Yellow Sea",
      Location == "Launch Plateform, Shahrud Missile Test Site" ~ "Shahrud Missile Test Site",
      Location == "Rocket Lab LC-1A, M?Ã\u0081hia Peninsula, New Zealand" ~ "New Zealand",
      TRUE ~ word(Location, -1)
    )
  )

```

```

Space_missions <- Space_missions %>%
  mutate(
    country = str_replace(country, "StatusRetired", replacement = "USA"),
    country = str_replace(country, "Yellow Sea", replacement = "China"),
    country = str_replace(country, "Russia", replacement = "Russian Federation"),
    country = str_replace(country, "Shahrud Missile Test Site", replacement = "Iran"),
    country = str_replace(country, "Range Facility", replacement = "USA"),
    country = str_replace(country, "Barents Sea", replacement = "Russia"),
    country = str_replace(country, "Canaria", replacement = "USA")
  )

```

```
Space_missions %>% count(country, sort = T)
```

	country	n
## 1	Russian Federation	1395
## 2	USA	1347
## 3	Kazakhstan	701
## 4	France	303
## 5	China	269
## 6	Japan	126
## 7	India	76
## 8	Pacific Ocean	36
## 9	Iran	14
## 10	New Zealand	13
## 11	Israel	11
## 12	Kenya	9
## 13	Korea	8
## 14	Australia	6
## 15	Mexico	4
## 16	Brazil	3
## 17	Russia	3

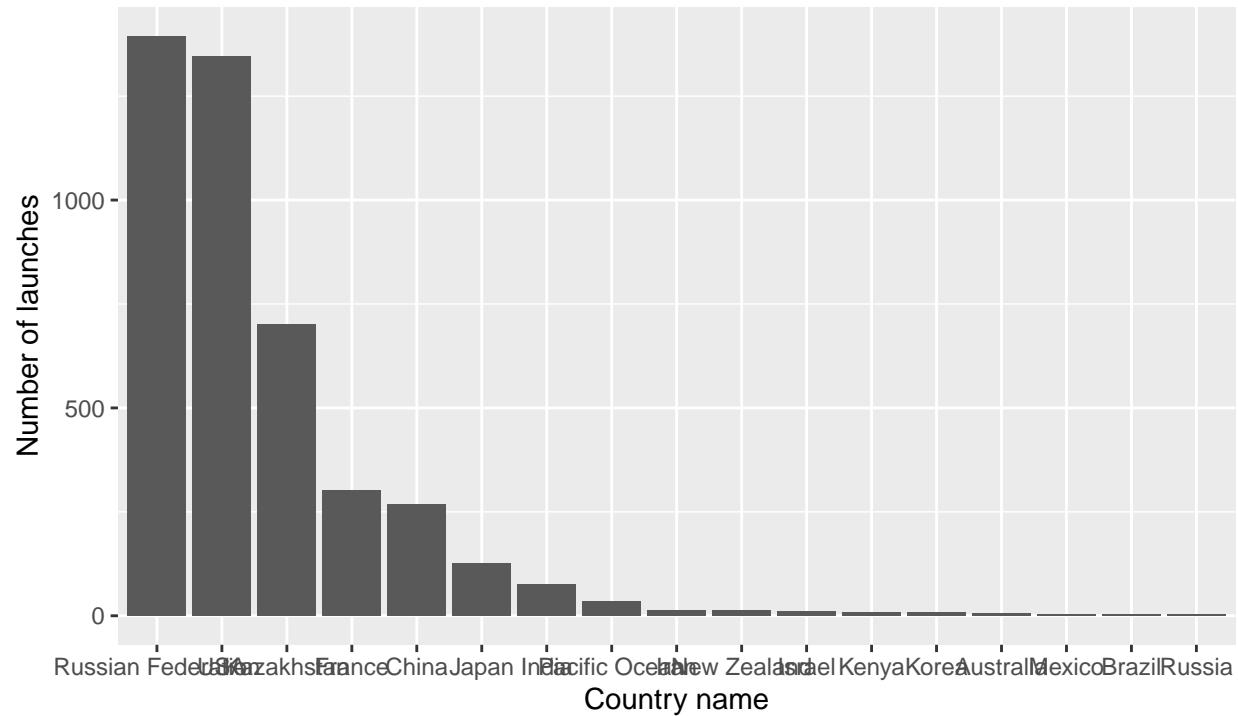
```

Space_missions %>%
  group_by(country) %>%
  summarise(count = n()) %>%
  group_by(country) %>%
  summarise(
    count_total = sum(count) # Calculate the total count for each country
  ) %>%
  arrange(desc(count_total)) %>% # Arrange countries by total count in descending order
  ggplot(aes(
    x = fct_reorder(country, -count_total), # Reorder countries by total number of launches
    y = count_total
  )) +
  geom_col() +
  labs(
    title = "Top 20 Companies",
    subtitle = "Column plot, Top 20 Companies",
    caption = "Kaggle: All Space Missions from 1957",
    x = "Country name",
    y = "Number of launches"
  )

```

Top 20 Companies

Column plot, Top 20 Companies



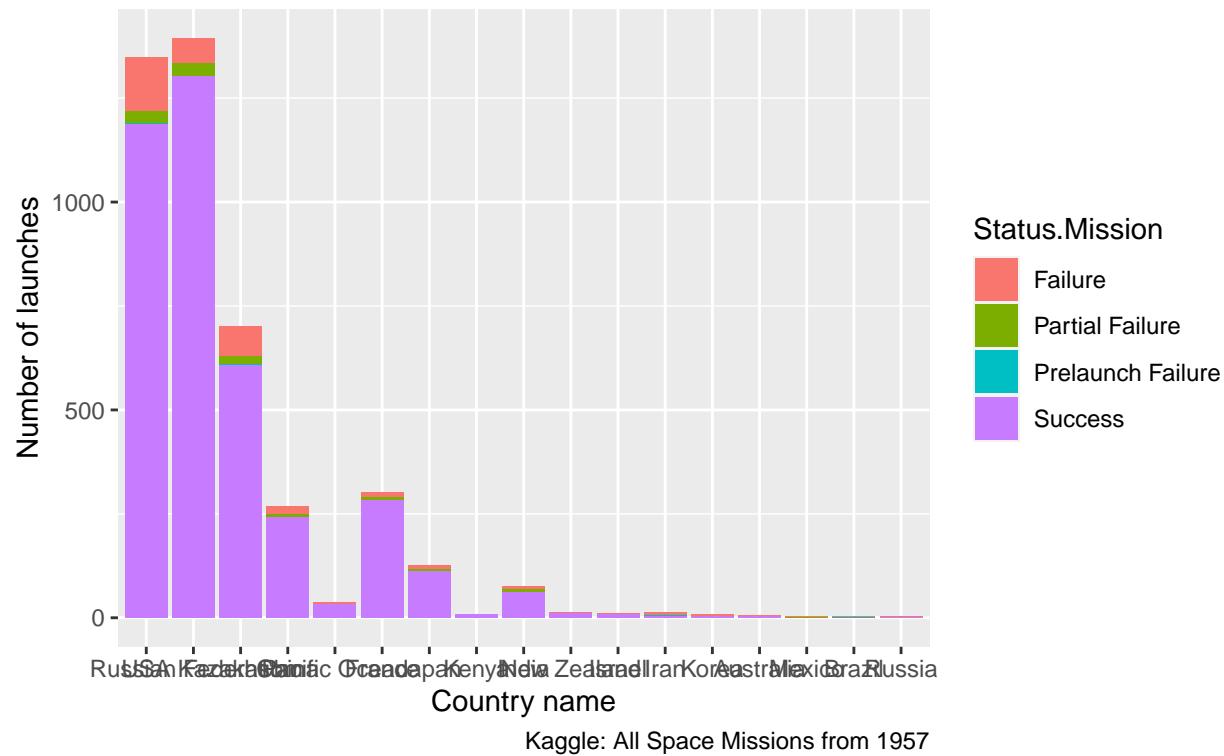
Kaggle: All Space Missions from 1957

##Plot

```
Space_missions %>%
  select(country, Status.Mission) %>%
  group_by(country, Status.Mission) %>%
  summarise(count = n()) %>%
  arrange(desc(count)) %>%
  ggplot(aes(x = fct_reorder(country, -count), y = count, fill = Status.Mission)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Top 20 Companies",
    subtitle = "Column plot, Top 20 Companies",
    caption = "Kaggle: All Space Missions from 1957",
    x = "Country name",
    y = "Number of launches"
  )
```

Top 20 Companies

Column plot, Top 20 Companies



##Tabulate missions for each country

```
Space_missions %>%
  count(country, Company.Name, sort = TRUE) %>%
  arrange(desc(n))
```

```
##           country   Company.Name     n
## 1  Russian Federation      RVSN USSR 1198
## 2       Kazakhstan      RVSN USSR  579
## 3         France      Arianespace 277
## 4         China        CASC 251
## 5         USA General Dynamics 251
## 6         USA          NASA 203
## 7         USA US Air Force 161
## 8  Russian Federation      VKS RF 157
## 9         USA            ULA 140
## 10        USA            Boeing 136
## 11        USA Martin Marietta 114
## 12        USA           SpaceX 100
## 13         Japan           MHI  84
## 14         USA Northrop  83
## 15         USA    Lockheed  79
## 16         India          ISRO  76
## 17  Kazakhstan      Roscosmos  47
## 18  Kazakhstan      VKS RF  44
## 19         USA           ILS  40
```

## 20	Pacific Ocean	Sea Launch	36
## 21	Japan	ISAS	30
## 22	USA	US Navy	17
## 23	France	ESA	13
## 24	Iran	ISA	13
## 25	New Zealand	Rocket Lab	13
## 26	Russian Federation	Eurockot	13
## 27	Kazakhstan	Kosmotras	12
## 28	USA	Blue Origin	12
## 29	Israel	IAI	11
## 30	China	ExPace	10
## 31	Russian Federation	Kosmotras	10
## 32	Kenya	ASI	9
## 33	France	CNES	8
## 34	Russian Federation	Roscosmos	8
## 35	Japan	JAXA	7
## 36	Kazakhstan	Land Launch	7
## 37	Russian Federation	MITT	7
## 38	USA	AMBA	7
## 39	Kazakhstan	ILS	6
## 40	China	CASIC	5
## 41	Japan	UT	5
## 42	Korea	KCST	5
## 43	France	Arm??e de l'Air	4
## 44	Mexico	Exos	4
## 45	Australia	CECLES	3
## 46	Brazil	AEB	3
## 47	Korea	KARI	3
## 48	Russia	SRC	3
## 49	Australia	RAE	2
## 50	Kazakhstan	Arianespace	2
## 51	Kazakhstan	OKB-586	2
## 52	Australia	AMBA	1
## 53	China	i-Space	1
## 54	China	Landspace	1
## 55	China	OneSpace	1
## 56	France	CECLES	1
## 57	Iran	IRGC	1
## 58	Kazakhstan	Starsem	1
## 59	Kazakhstan	Yuzhmash	1
## 60	Russian Federation	Khrunichev	1
## 61	Russian Federation	Yuzhmash	1
## 62	USA	Douglas	1
## 63	USA	EER	1
## 64	USA	Sandia	1
## 65	USA	Virgin Orbit	1

```
Space_missions %>%
  filter(country == "Russian Federation") %>%
  count(Company.Name, sort = TRUE) %>%
  arrange(desc(n))
```

```
##   Company.Name    n
## 1     RVSN USSR 1198
```

```

## 2      VKS RF  157
## 3    Eurockot   13
## 4    Kosmotras   10
## 5    Roscosmos    8
## 6        MITT    7
## 7  Khrunichev    1
## 8    Yuzhmash    1

Space_missions %>%
  filter(country == "USA") %>%
  count(Company.Name, sort = TRUE) %>%
  arrange(desc(n))

```

```

##           Company.Name   n
## 1  General Dynamics 251
## 2            NASA 203
## 3  US Air Force 161
## 4          ULA 140
## 5        Boeing 136
## 6 Martin Marietta 114
## 7       SpaceX 100
## 8  Northrop 83
## 9    Lockheed 79
## 10       ILS 40
## 11    US Navy 17
## 12  Blue Origin 12
## 13     AMBA 7
## 14     Douglas 1
## 15      EER 1
## 16     Sandia 1
## 17 Virgin Orbit 1

```

```

Space_missions %>%
  filter(country == "China") %>%
  count(Company.Name, sort = TRUE) %>%
  arrange(desc(n))

```

```

##           Company.Name   n
## 1          CASC 251
## 2        ExPace 10
## 3        CASIC 5
## 4        i-Space 1
## 5      Landspace 1
## 6      OneSpace 1

```

##Plot

```

Space_missions %>%
  group_by(country, Company.Name) %>%
  summarise(n = n()) %>%
  group_by(country) %>%
  summarise(total = sum(n)) %>%

```

```

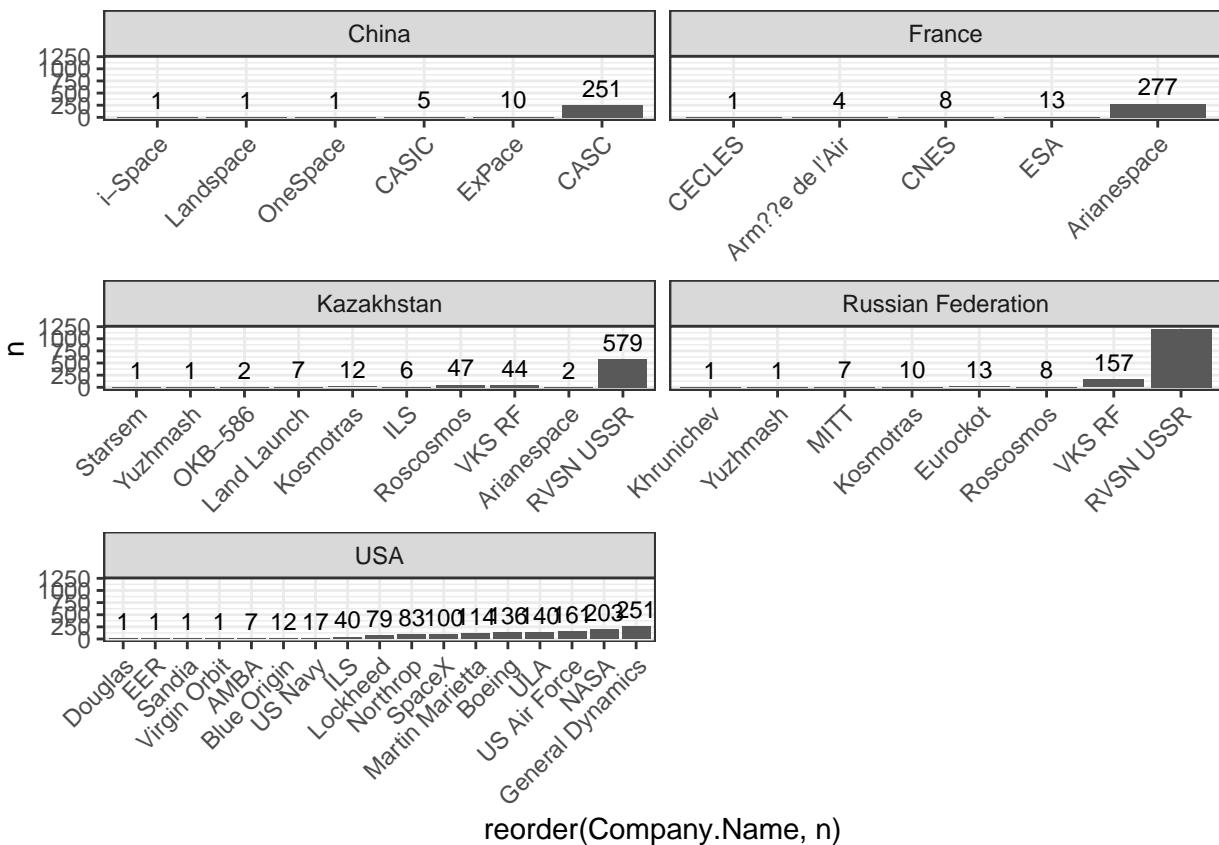
top_n(5, total) %>%
inner_join(Space_missions, by = "country") %>%
group_by(country, Company.Name) %>%
summarise(n = n()) %>%
ggplot(aes(x = reorder(Company.Name, n), y = n)) +
geom_col() +
geom_text(aes(label = n), vjust = -0.5, size = 3) +
theme_bw() +
facet_wrap(~country, ncol = 2, scales = "free_x") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

```

## `summarise()` has grouped output by 'country'. You can override using the
## '.groups' argument.
## `summarise()` has grouped output by 'country'. You can override using the
## '.groups' argument.

```



```

##Plot for the big three nations

```

```

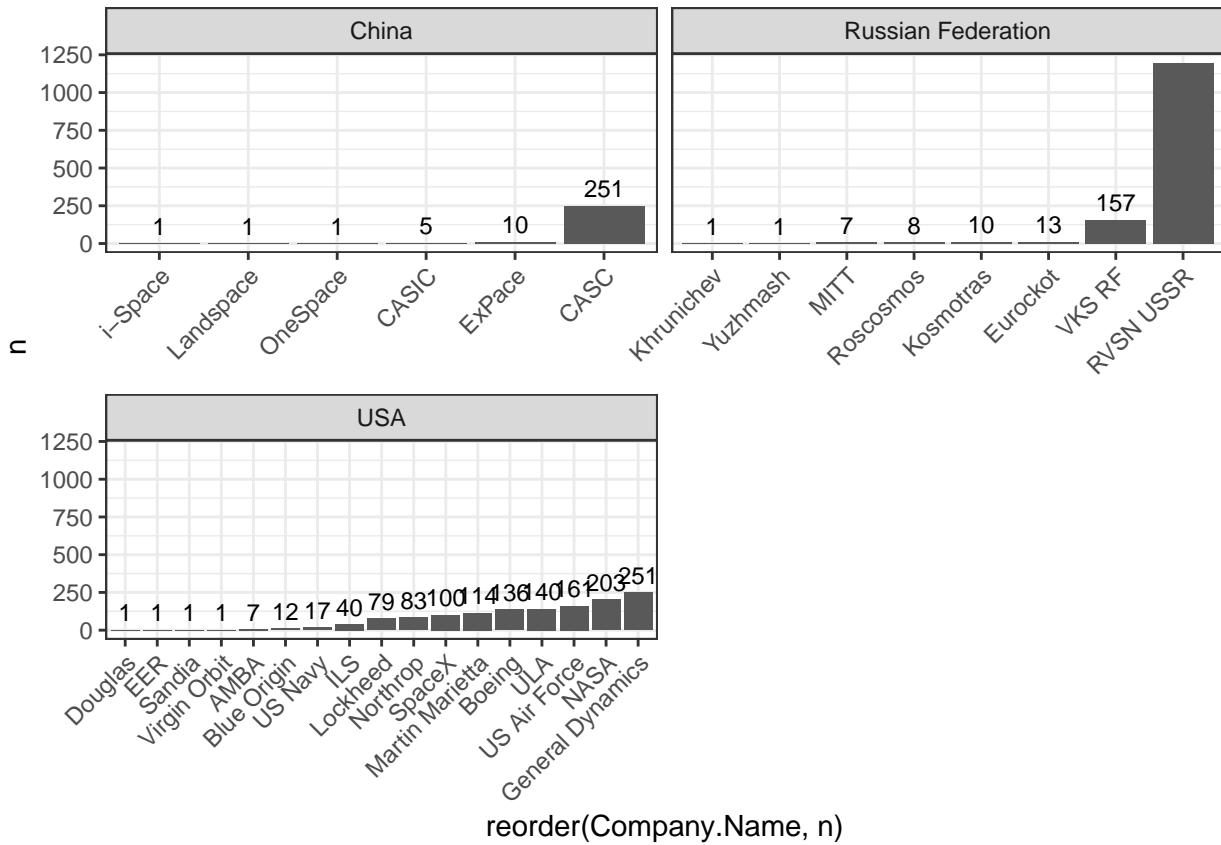
Space_missions %>%
filter(country %in% c("USA", "Russian Federation", "China")) %>%
group_by(country, Company.Name) %>%
summarise(n = n()) %>%
group_by(country) %>%
summarise(total = sum(n)) %>%
top_n(5, total) %>%

```

```

inner_join(Space_missions, by = "country") %>%
group_by(country, Company.Name) %>%
summarise(n = n()) %>%
ggplot(aes(x = reorder(Company.Name, n), y = n)) +
geom_col() +
geom_text(aes(label = n), vjust = -0.5, size = 3) +
theme_bw() +
facet_wrap(~country, ncol = 2, scales = "free_x") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



Plot in red the column which will be used for the graphic

```

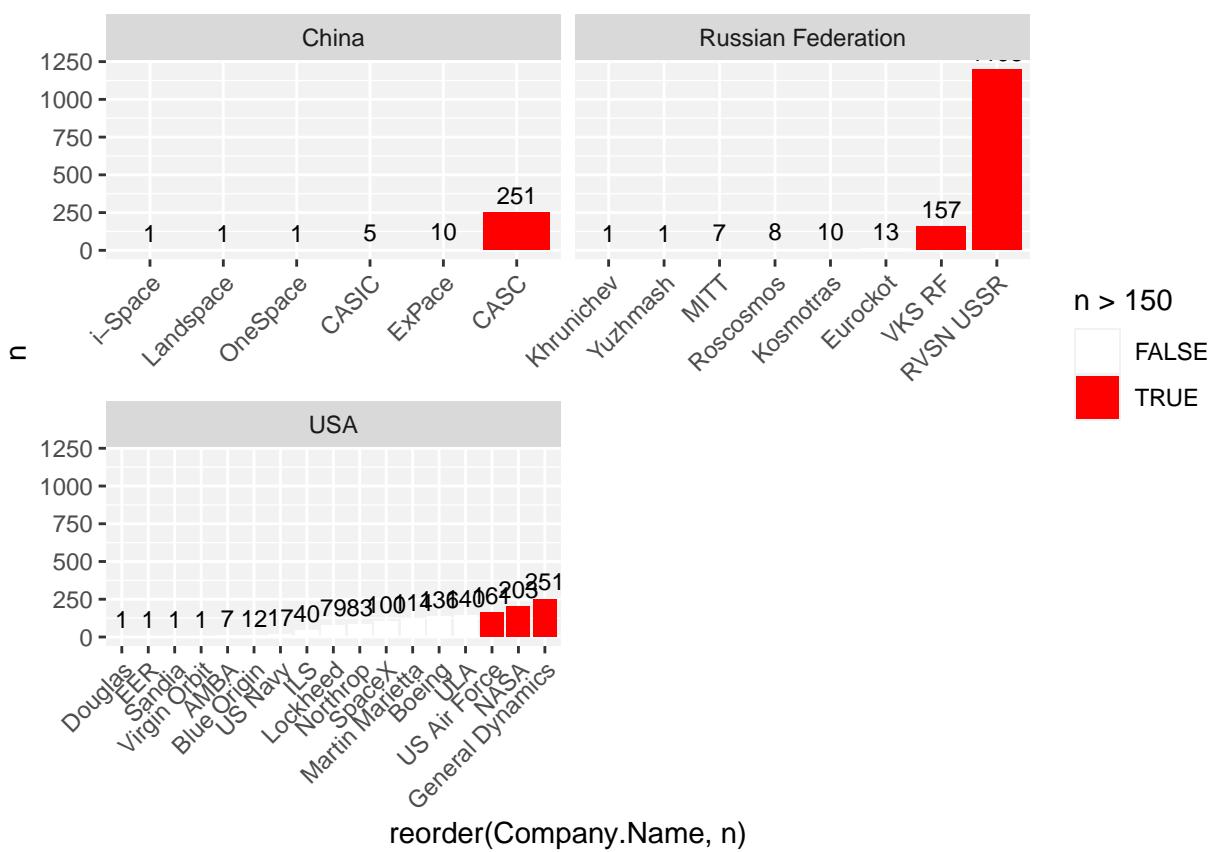
Space_missions %>%
filter(country %in% c("USA", "Russian Federation", "China")) %>%
group_by(country, Company.Name) %>%
summarise(n = n()) %>%
group_by(country) %>%
summarise(total = sum(n)) %>%
top_n(5, total) %>%
inner_join(Space_missions, by = "country") %>%
group_by(country, Company.Name) %>%
summarise(n = n()) %>%
ggplot(aes(x = reorder(Company.Name, n), y = n, fill = n > 150)) +

```

```

geom_col() +
geom_text(aes(label = n), vjust = -0.5, size = 3) +
facet_wrap(~country, ncol = 2, scales = "free_x") +
scale_fill_manual(values = c("FALSE" = "white", "TRUE" = "red")) +
theme(
  panel.background = element_rect(fill = "gray95", colour = NA)
) +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



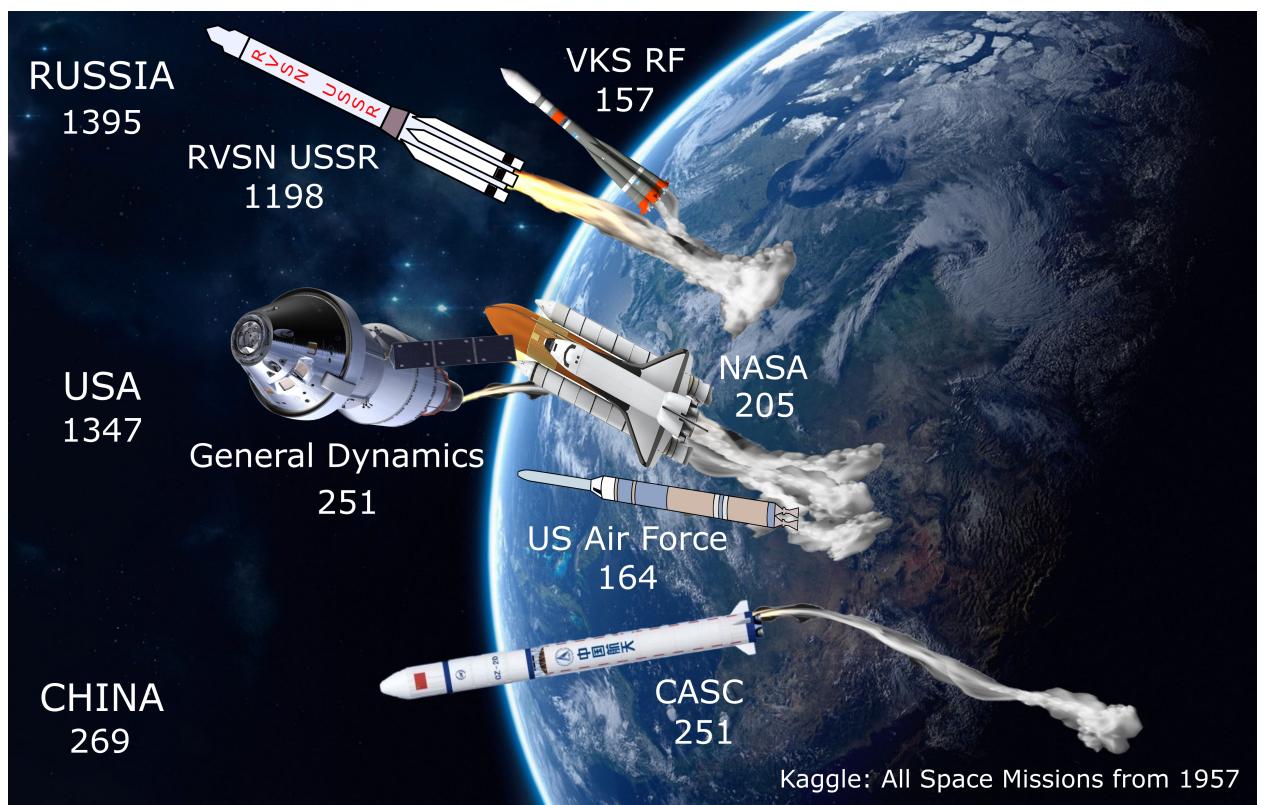


Figure 1: Quick illustration and final image