Package and Repository Analysis

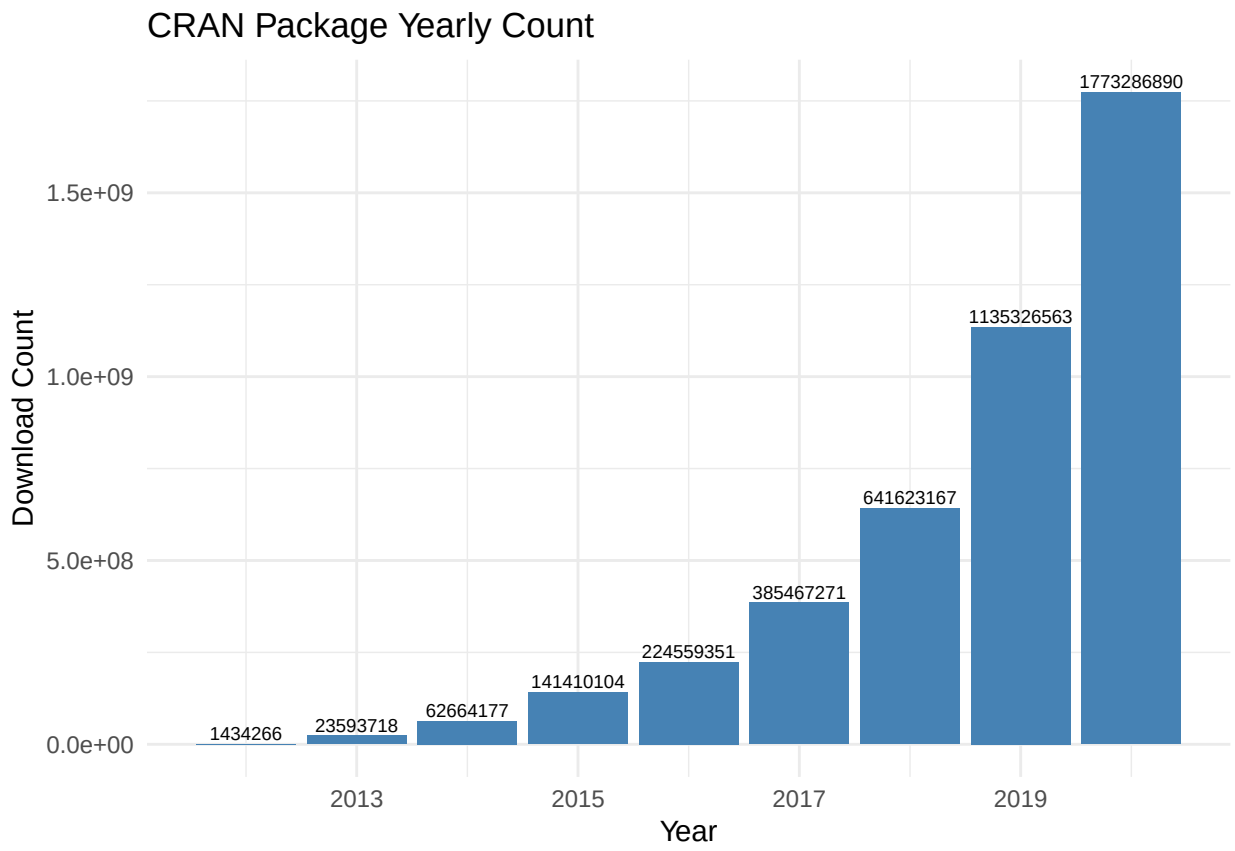
Meet Bhatnagar

16/08/2021

CRAN Analysis

1. Yearly chart of CRAN package downloads

```
library(jsonlite)
library(ggplot2)
library(lubridate)
library(dplyr)
dailyPKG = fromJSON("https://cranlogs.r-pkg.org/downloads/daily/2012-01-01:2020-12-31")
downloads <- dailyPKG$downloads[[1]]
dailypkg_dwnld <- data.frame(Date=as.Date(downloads$day), Count=downloads$downloads)
dailypkg_dwnld$Year = year(dailypkg_dwnld$Date)
yearly_dwnld <- dailypkg_dwnld %>%
  group_by(Year) %>%
  summarise(dwnld_count = sum(Count))
ggplot(data=yearly_dwnld, aes(x=Year, y=dwnld_count)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=dwnld_count), vjust=-0.3, size=2.6)+
  labs(title="CRAN Package Yearly Count", x="Year", y = "Download Count")+
  theme_minimal()
```



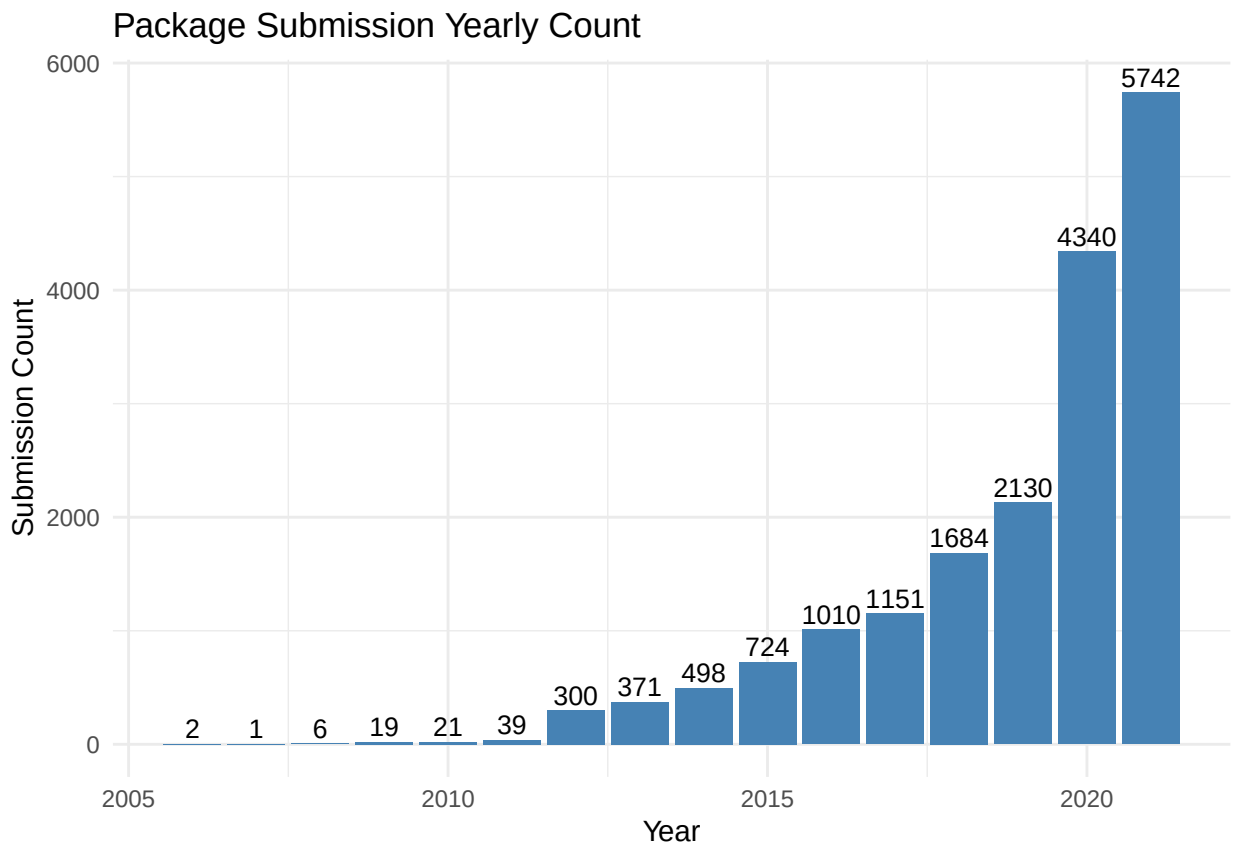
2. Yearly chart of CRAN package submissions

```
library(rvest)
library(xml2)
library(lubridate)
CRAN_package <- read_html("https://cran.r-project.org/web/packages/available_packages_by_date.html")

pkg_subm_date <- CRAN_package %>%
  html_nodes("td:nth-child(1)") %>%
  html_text()
pkg_subm_date <- trimws(pkg_subm_date)

pkg_subm_df <- data.frame(Year = year(as.Date(pkg_subm_date)))
pkg_subm_df <- pkg_subm_df %>%
  group_by(Year) %>%
  summarise(subm_count = n())

ggplot(data=pkg_subm_df, aes(x=Year, y=subm_count)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=subm_count), vjust=-0.3, size=3.6)+
  labs(title="Package Submission Yearly Count", x="Year", y = "Submission Count")+
  theme_minimal()
```

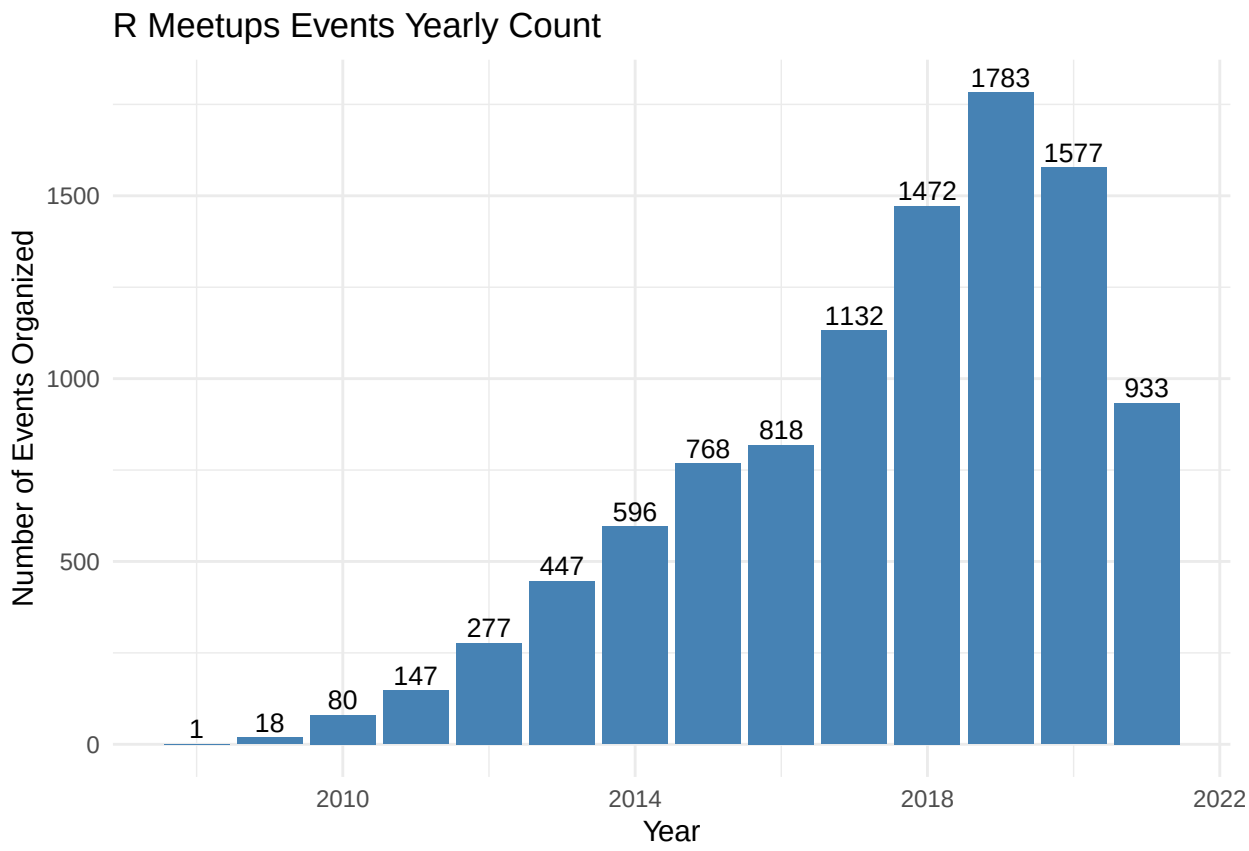


R meetup events

Yearly chart of R meetup events

```
past_event <- readRDS("all_past_R_events.rds")
past_event$Year = year(past_event$local_date)
yearly_event_count <- past_event %>%
  group_by(Year) %>%
  summarise(Count = n())

ggplot(data=yearly_event_count, aes(x=Year, y=Count)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=Count), vjust=-0.3, size=3.6)+
  labs(title="R Meetups Events Yearly Count", x="Year", y = "Number of Events Organized")+
  theme_minimal()
```

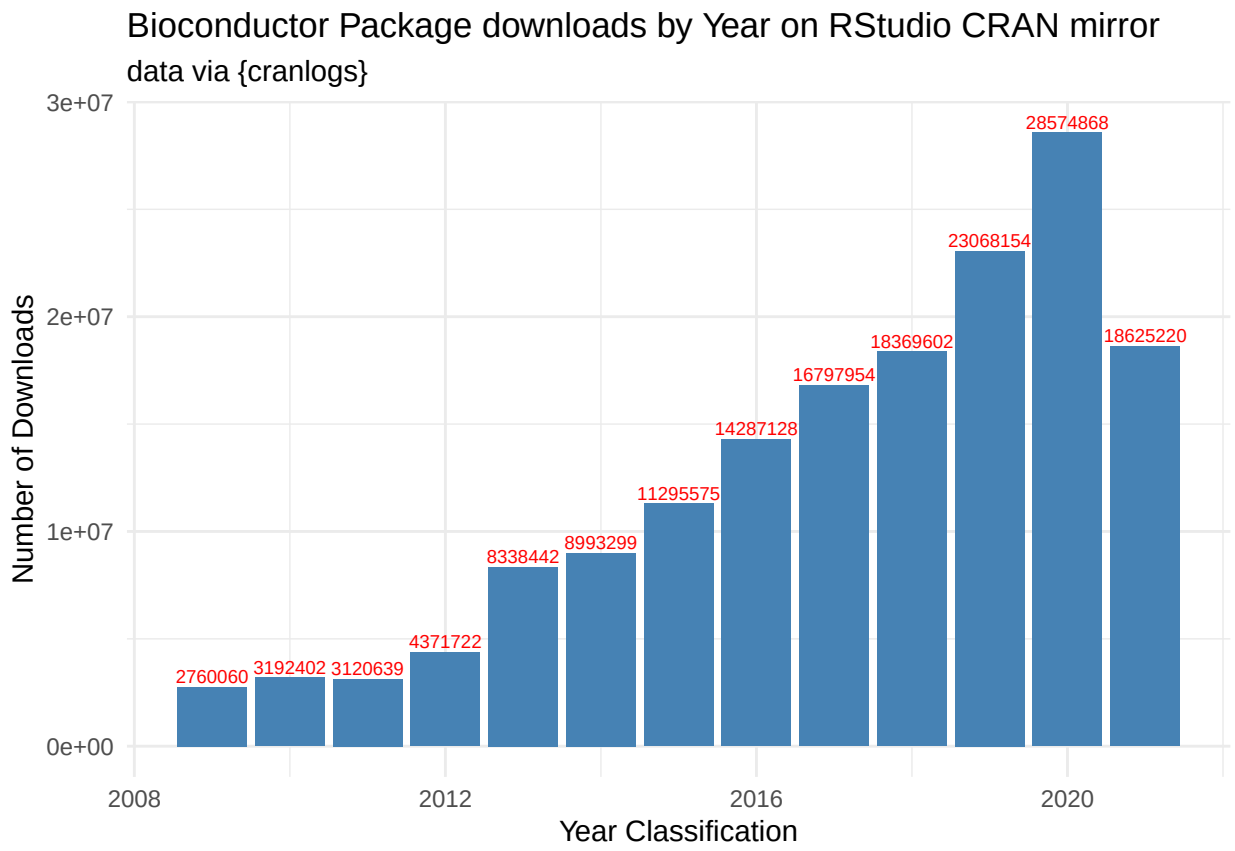


Bioconductor

Yearly chart of Bioconductor package downloads

```
library(data.table)
library(packageRank)
library(magrittr)
library(ggplot2)

bioc_dwnld_df <- bioconductorDownloads()
bioc_dwnld_df <- bioc_dwnld_df[["data"]][[1]]
setDT(bioc_dwnld_df)
bioc_dwnld_df[, .(count = sum(Nb_of_downloads)), Year] %>%
  ggplot(aes(Year, count)) +
  geom_bar(stat="identity", fill="steelblue") +
  geom_text(aes(label=count), vjust=-0.3, color="red", size=2.3) +
  labs(
    title = "Bioconductor Package downloads by Year on RStudio CRAN mirror",
    subtitle = "data via {cranlogs}",
    x = "Year Classification", y="Number of Downloads"
  ) + theme_minimal()
```



GitHub

Yearly chart of R GitHub package repository

```
library(httr)
library(jsonlite)
library(lubridate)
library(ggplot2)

year_repo_count <- c()
timeline <- 2008:year(Sys.Date())
for(i in 1:length(timeline)){
  url <- "https://api.github.com"
  path <- paste0("search/repositories?q=language:R+created:",as.character(timeline[i]),"&per_page=100&s
  raw.result <- GET(url = url, path = path)
  this.raw.content <- rawToChar(raw.result$content)
  this.content <- fromJSON(this.raw.content)
  year_repo_count[i] <- this.content$total_count
  Sys.sleep(5)
}
yearly_repo_df <- data.frame(Year = 2008 : year(Sys.Date()),
                             RepoCount = year_repo_count)
ggplot(data=yearly_repo_df, aes(x=Year, y=RepoCount)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=RepoCount), vjust=-0.3, color="black", size=3.5)+
  labs(
```

```
  title = "Count of R Repositories on GitHub - Yearly Basis",
  x = "Year", y="Number of R Repositories"
) + theme_minimal()
```