R Notebook

```
stages = rbind(stagesE1, stagesE2, stagesE3) %>% select(Interval, Stage, Executors, Duration) %>%
   mutate(Stage = paste(str_pad(Interval,2,"left"), Stage)) %>%
   select(Stage, Executors, Duration) %>%
   group_by(Stage, Executors) %>% summarise(Duration = mean(Duration))
head(stages)
## # A tibble: 6 x 3
## # Groups:
                      Stage [2]
                                               Executors Duration
##
       Stage
##
       <chr>
                                               <chr>
                                                                   <dbl>
## 1 " 0 1.Maximal disks found" 12
                                                                    39.2
                                                                    35.8
## 2 " 0 1.Maximal disks found" 4
## 3 " 0 1.Maximal disks found" 8
                                                                    35.8
## 4 " 1 1.Maximal disks found" 12
                                                                    42.2
## 5 " 1 1.Maximal disks found" 4
                                                                    28.6
## 6 " 1 1.Maximal disks found" 8
                                                                    29.7
p = ggplot(data = stages, aes(x = Stage, y = Duration, fill = Executors)) +
   geom_bar(stat="identity", position=position_dodge(width = 0.75), width = 0.7) +
   theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
   labs(x="Stages", y="Duration(s)")
plot(p)
    50 -
    40
                                                                                                                  Executors
Duration(s)
    30
                                                                                                                         12
                                                                                                                         4
    20
                                                                                                                         8
    10
                                                                                                     3 5.Flocks reported T
                                                                                                          3 6.Candidates indexed -
           0 1.Maximal disks found
                1 1.Maximal disks found -
                     1 2.Join done 7
                                                     2 2.Join done
                                                          2 3.Flocks to report 7
                                                                2 4. Flocks prunned by expansion
                                                                          2 6.Candidates indexed -
                                                                                3 1.Maximal disks found
                                                                                          3 3.Flocks to report -
                                                                                               3 4.Flocks prunned by expansion
                                           1 6.Candidates indexed -
                                                2 1.Maximal disks found
                                     1 5.Flocks reported
                                                                                     3 2. Join done
                                                                     2 5.Flocks reported
                           1 3.Flocks to report
                                1 4.Flocks prunned by expansion
                                                      Stages
```

```
stages = customExecutionTime(nohup)
data = stages %>% select(Executors, Epsilon, Duration) %>%
  group_by(Executors, Epsilon) %>% summarise(Duration = mean(Duration))
head(data)
## # A tibble: 3 x 3
## # Groups: Executors [3]
    Executors Epsilon Duration
##
     <chr>
               <chr>
                           <dbl>
## 1 12
               110.0
                            329.
## 2 4
               110.0
                            275.
## 3 8
               110.0
                            280.
   300 -
Duration(s)
                                                                                Executors
                                                                                    12
                                                                                    4
                                                                                    8
   100 -
     0 -
                                   Epsilon(mts)
```