

Untitled

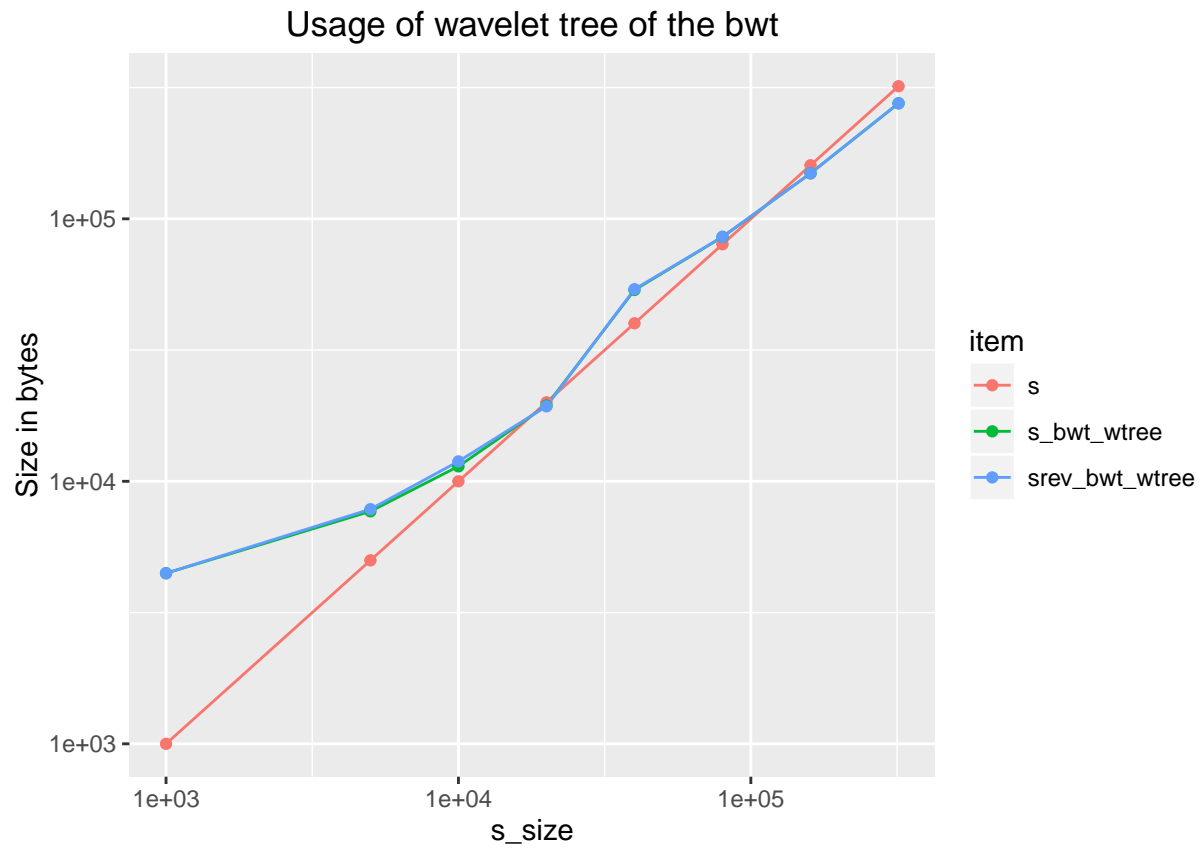
O. Denas

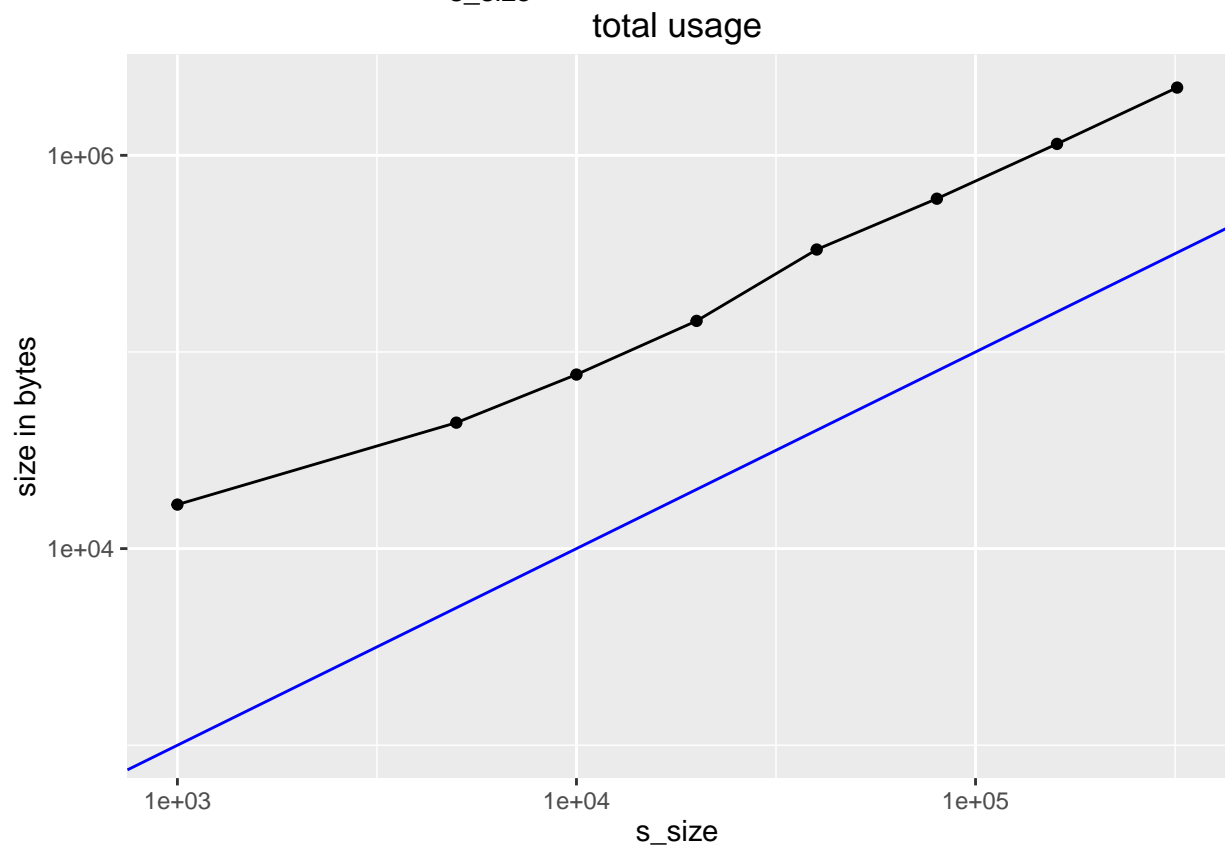
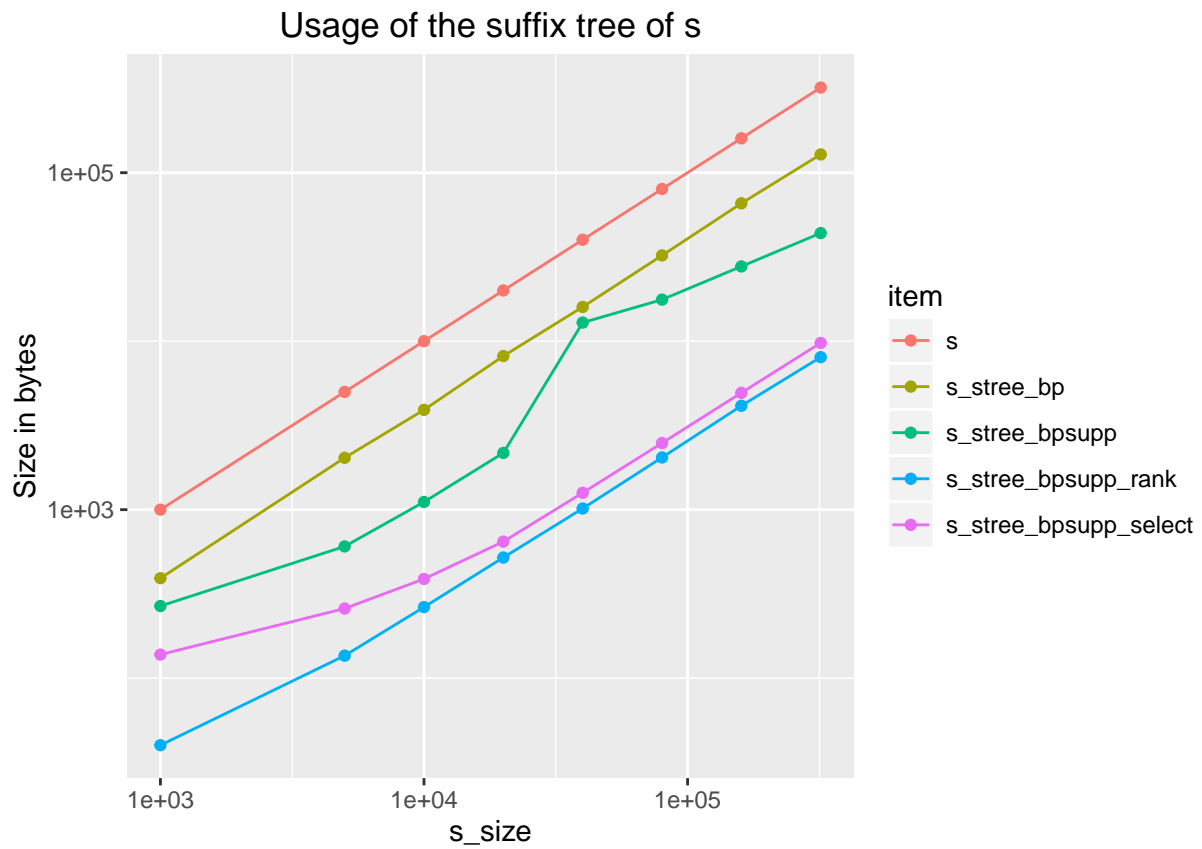
November 20, 2016

R Markdown

Testing space usage on subsets of the `proteins.50MB` dataset from Pizza&Chilli.

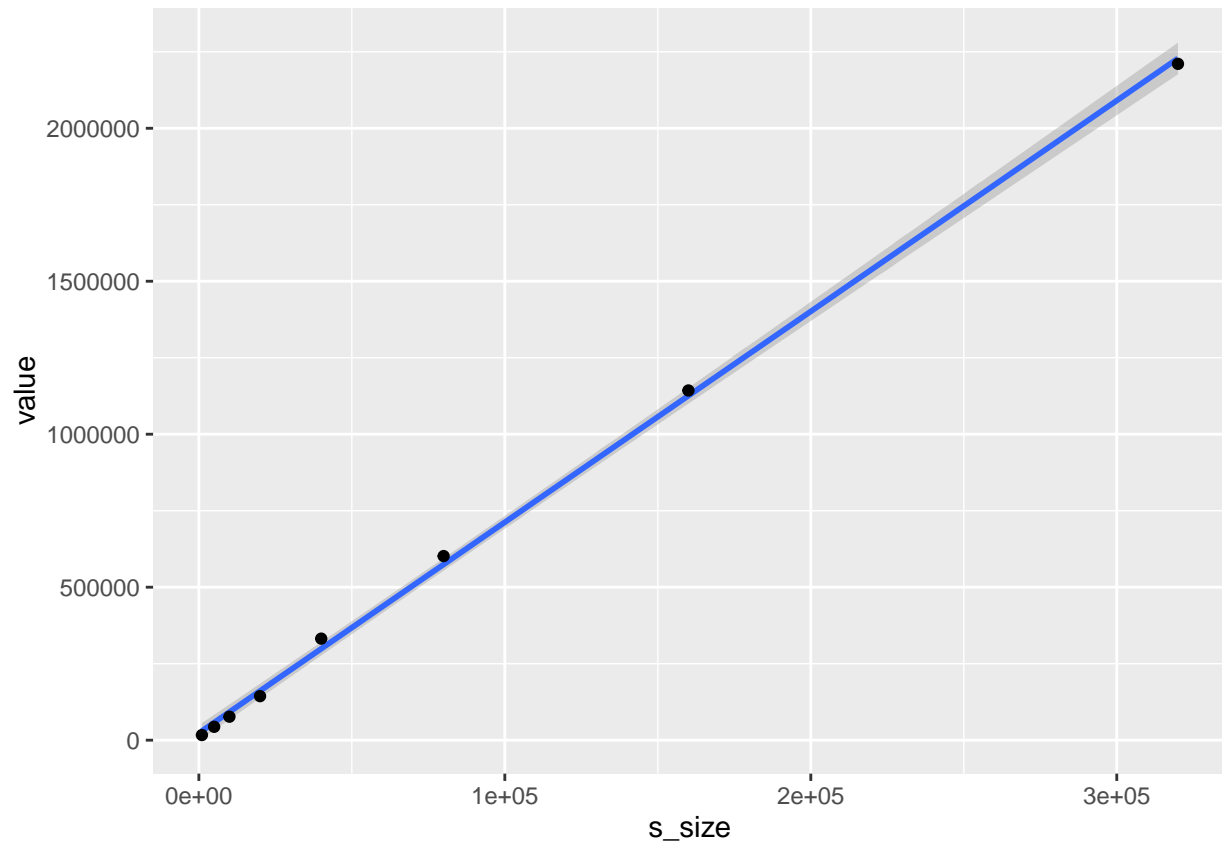
Below is space usage in bytes as a function of length of `s`. All datasets had a text `t` of length 1000.





Fit a line to the points ...

```
ggplot(raw_ds %>% group_by(prefix) %>% summarise(value=sum(value), s_size=mean(s_size)), aes(s_size, value))
```



```
lm(value ~ s_size, data=raw_ds %>% group_by(prefix) %>% summarise(value=sum(value), s_size=mean(s_size)))
```

```
##
## Call:
## lm(formula = value ~ s_size, data = raw_ds %>% group_by(prefix) %>%
##   summarise(value = sum(value), s_size = mean(s_size)))
##
## Coefficients:
## (Intercept)      s_size
##  23203.834      6.891
```