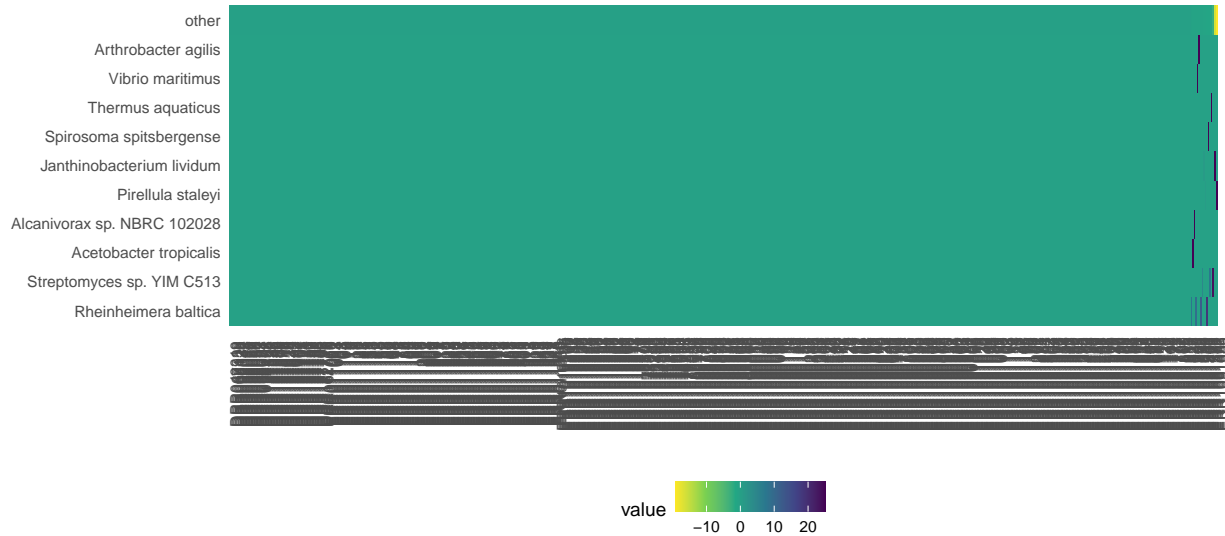# Basic samovar usage

Download data

Download and vizualize data from GMrepo

```
teatree <- GMrepo_type2data(number_to_process = 2500)
viz_composition(teatree, type = "tile", interactive = F, top = 10)
```


Preprocessing

Filter data

```
tealeaves <- teatree %>%
  teatree_trim(treshhold_species = 3,
               treshhold_samples = 3,
               treshhold_amount = 10^(-3))
#> Initial: 3031 x 632
#> After filtering: 482 x 621
```
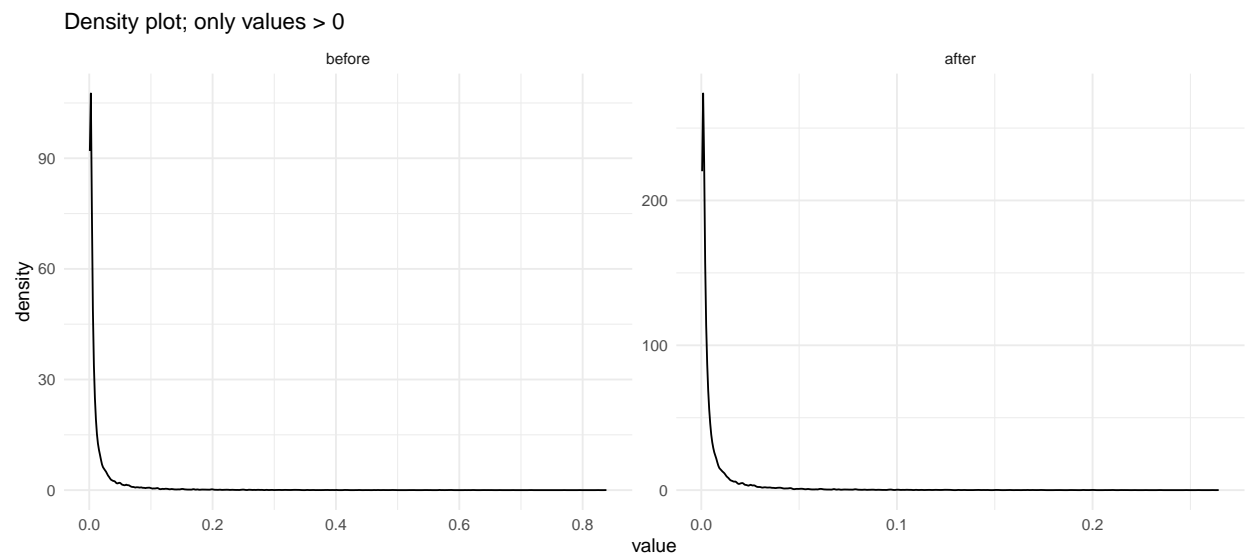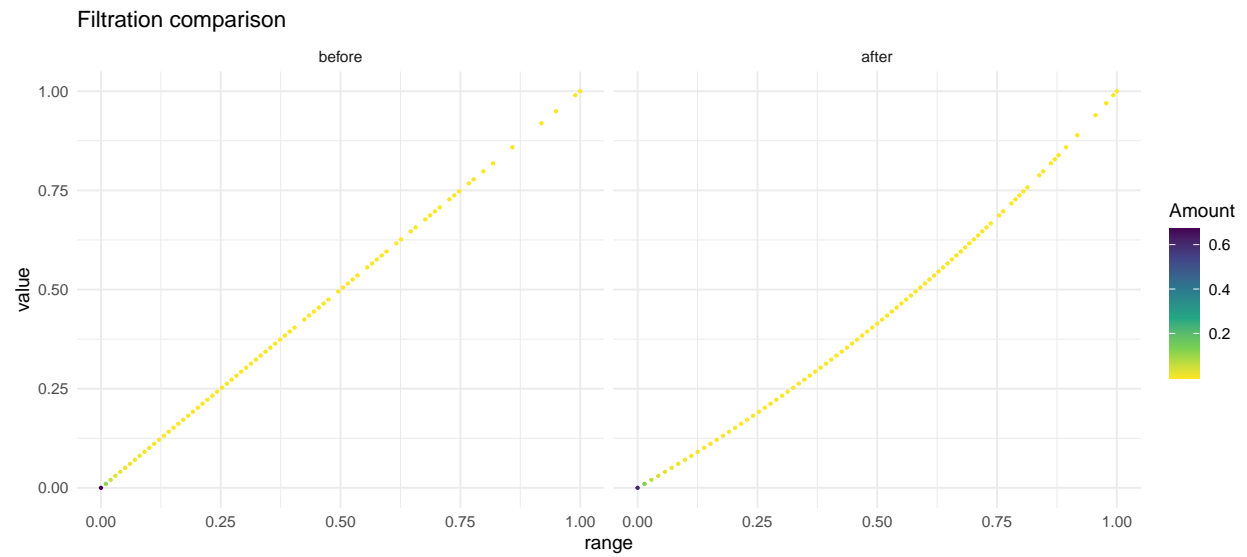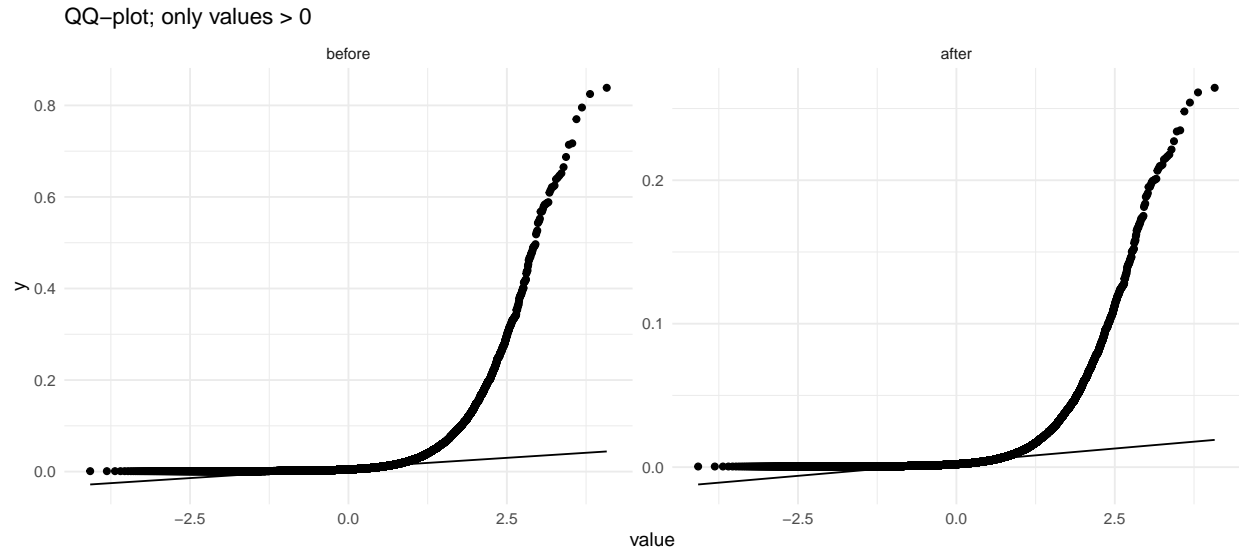
Normalizing

If you build teatree by your own, do rescaling stage when building via `teatree$rescale()` or assigning `teatree$min_value` and `teatree$max_value` is required

Good approximation to normal distribution is required for `glm` generating methods

```
teabag <- tealeaves %>%
  tealeaves_pack()
```

Filtration comparison

Density plot; only values > 0

QQ–plot; only values > 0

Clustering

Do hierarchical (deprecated) or aggregating clustering of species

Remember: if you want to re-filter, it is better to re-do welding stage to avoid crashes in future!

```
concotion <- teabag %>%
  teabag_brew(min_cluster_size = 40, max_cluster_size = 150)
#> Warning in reducing maximum sample: to similar species!
#> If you want to use bootstrap for samovaR build, it might be better to re-filter species
```
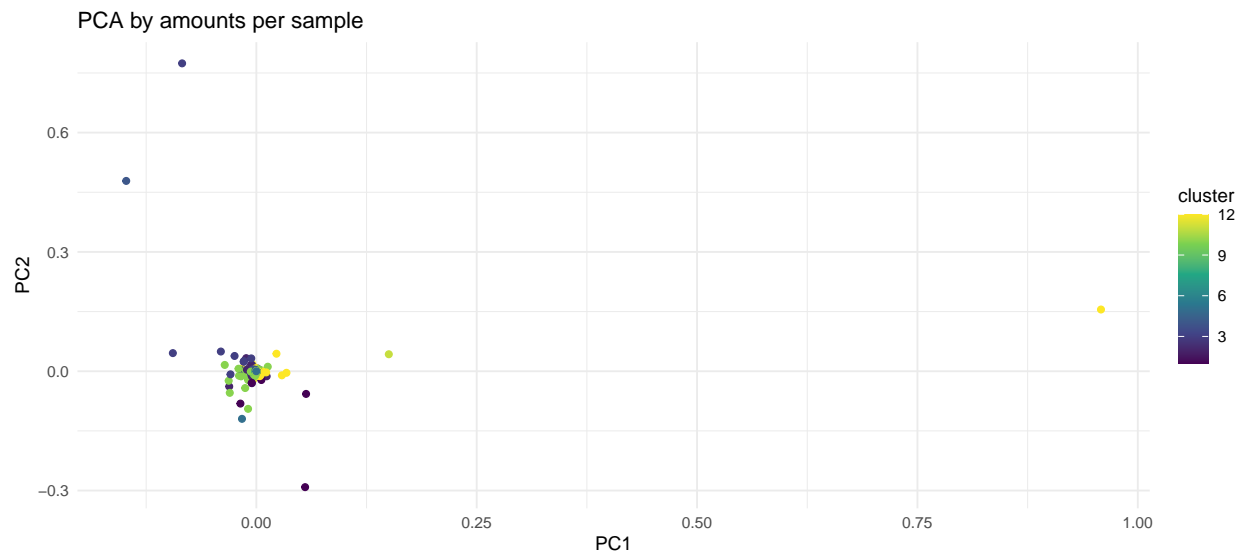

Build samovar

```
data.samovar <- concotion %>%
  concotion_pour(probability_calculation = "simple")
#> ---  Prepare tea leaves for different clusters  ---
#>
```

```
#> --- Prepare tea leaves for calculation between clusters ---
#>
#> --- Samovar built ---
```

Generate data

```
new_data <- data.samovar %>%
  samovar_boil(N = 100, avoid_zero_generations = T)
#> --- Generation started ---
#>
#> --- Generation done ---
```

The generated data:

```
viz_composition(new_data, reord_samples = "hcl",type = "tile", interactive = F, top = 10)
```