R Notebook

```
library(ggplot2)
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(patchwork)
library(ggraph)
library(tidygraph)
```

```
##
## Attaching package: 'tidygraph'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(ggthemes)
library(RColorBrewer)
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following object is masked from 'package:tidygraph':
##
##     groups
```

```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following object is masked from 'package:tidyr':
##
##     crossing
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
villagers <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/
```

```
## Parsed with column specification:
## cols(
##   row_n = col_double(),
##   id = col_character(),
##   name = col_character(),
##   gender = col_character(),
##   species = col_character(),
##   birthday = col_character(),
##   personality = col_character(),
##   song = col_character(),
##   phrase = col_character(),
##   full_id = col_character(),
##   url = col_character()
## )
```

goal: draw a sankey diagram to show how gender, species and personality are connected

```
# need to set up dataframe with sources and targets
# levels gender --> personality --> species
head(villagers)
```

```
## # A tibble: 6 x 11
##   row_n id    name  gender species birthday personality song  phrase full_id
##   <dbl> <chr> <chr> <chr>  <chr>   <chr>    <chr>       <chr> <chr>  <chr>
## 1     2 admi~ Admi~ male   bird    1-27     cranky      Stee~ aye a~ villag~
## 2     3 agen~ Agen~ female squirr~ 7-2      peppy       DJ K~ sidek~ villag~
## 3     4 agnes Agnes female pig     4-21     uchi        K.K.~ snuff~ villag~
## 4     6 al    Al    male   gorilla 10-18    lazy        Stee~ Ayyee~ villag~
## 5     7 alfo~ Alfo~ male   alliga~ 6-9      lazy        Fore~ it'sa~ villag~
## 6     8 alice Alice female koala   8-19     normal      Surf~ guvnor villag~
## # ... with 1 more variable: url <chr>
```

```
unique(villagers$species)
```

```
##  [1] "bird"      "squirrel"  "pig"       "gorilla"   "alligator" "koala"
##  [7] "eagle"     "anteater"  "bull"      "mouse"     "cat"       "horse"
## [13] "hamster"   "kangaroo"  "wolf"      "penguin"   "chicken"   "elephant"
## [19] "sheep"     "deer"      "tiger"     "cub"       "dog"       "bear"
## [25] "hippo"     "duck"      "goat"      "ostrich"   "rabbit"    "lion"
## [31] "frog"      "monkey"    "rhino"     "octopus"   "cow"
```

add in a new types of column with a category for type of species

```
villagers<-villagers %>%
          mutate(species_type='mammal') %>%
          mutate(species_type=ifelse(species %in% c('bird','eagle','penguin','chicken','duck','ostrich'
          mutate(species_type=ifelse(species %in% c('octopus'),yes='invert',no=species_type)) %>%
         mutate(species_type=ifelse(species %in% c('alligator'),yes='reptile',no=species_type)) %>%
        mutate(species_type=ifelse(species %in% c('frog'),yes='amphibian',no=species_type))
head(villagers)
```

```
## # A tibble: 6 x 12
##   row_n id    name  gender species birthday personality song  phrase full_id
##   <dbl> <chr> <chr> <chr>  <chr>   <chr>    <chr>       <chr> <chr>  <chr>
## 1     2 admi~ Admi~ male   bird    1-27     cranky      Stee~ aye a~ villag~
```

```
## 2     3 agen~ Agen~ female squirr~ 7-2      peppy       DJ K~ sidek~ villag~
## 3     4 agnes Agnes female pig     4-21     uchi        K.K.~ snuff~ villag~
## 4     6 al    Al    male   gorilla 10-18    lazy        Stee~ Ayyee~ villag~
## 5     7 alfo~ Alfo~ male   alliga~ 6-9      lazy        Fore~ it'sa~ villag~
## 6     8 alice Alice female koala   8-19     normal      Surf~ guvnor villag~
## # ... with 2 more variables: url <chr>, species_type <chr>
```

create graph

```
#set up node levels
node_0<- villagers %>%
          mutate(name='villagers') %>%
          group_by(name) %>%
          summarize(size=n()) %>%
          mutate(shortname='villagers')  %>%
          mutate(type='root')

node_1<- villagers %>%
          mutate(name=paste('villagers',gender,sep='.')) %>%
          group_by(name) %>%
          summarize(size=n()) %>%
          mutate(shortname=gsub('villagers.','',name)) %>%
          mutate(type='gender')

node_2<- villagers %>%
          mutate(name=paste('villagers',gender,personality,sep='.')) %>%
          group_by(name) %>%
          summarize(size=n()) %>%
          mutate(shortname=gsub('villagers\\.[a-z]*\\.','',name)) %>%
          mutate(type='personality')

 node_3<- villagers %>%
          mutate(name=paste('villagers',gender,personality,species_type,sep='.')) %>%
          group_by(name) %>%
          summarize(size=n()) %>%
          mutate(temp=name) %>%
          separate(temp,into=c('rm1','rm2','rm3','shortname'),sep='\\.',) %>%
          select(name,size,shortname) %>%
          mutate(type='species')

nodes<-bind_rows(node_0,node_1,node_2,node_3)
head(nodes)
```

```
## # A tibble: 6 x 4
##   name                      size shortname type
##   <chr>                    <int> <chr>     <chr>
## 1 villagers                  391 villagers root
## 2 villagers.female           187 female    gender
## 3 villagers.male             204 male      gender
## 4 villagers.female.normal     59 normal    personality
## 5 villagers.female.peppy      49 peppy     personality
## 6 villagers.female.snooty     55 snooty    personality
```

Add node info for labels

```
nodes <- nodes %>%
        mutate(label='') %>%
        mutate(label=ifelse(type %in% c('root','gender','personality'),yes=shortname,no=label))
nodes
```

```
## # A tibble: 43 x 5
##    name                     size shortname type        label
##    <chr>                   <int> <chr>     <chr>       <chr>
##  1 villagers                 391 villagers root        villagers
##  2 villagers.female          187 female    gender      female
##  3 villagers.male            204 male      gender      male
##  4 villagers.female.normal    59 normal    personality normal
##  5 villagers.female.peppy     49 peppy     personality peppy
##  6 villagers.female.snooty    55 snooty    personality snooty
##  7 villagers.female.uchi      24 uchi      personality uchi
##  8 villagers.male.cranky      55 cranky    personality cranky
##  9 villagers.male.jock        55 jock      personality jock
## 10 villagers.male.lazy        60 lazy      personality lazy
## # ... with 33 more rows
```

Set up edges

```
#edges to and from directions
#gender to personality
#personality to gender

edges_0<-villagers %>%
          select(gender) %>%
          mutate(to='villagers')  %>%
          mutate(from=paste('villagers',gender,sep='.')) %>%
          unique() %>%
          select(to,from)

edges_1<-villagers %>%
          select(gender,personality) %>%
          mutate(to=paste('villagers',gender,sep='.'))  %>%
          mutate(from=paste('villagers',gender,personality,sep='.')) %>%
          unique() %>%
          select(to,from)

edges_2<-villagers %>%
          select(gender,personality,species_type) %>%
          mutate(to=paste('villagers',gender,personality,sep='.'))  %>%
          mutate(from=paste('villagers',gender,personality,species_type,sep='.')) %>%
          unique() %>%
          select(to,from)
edges<-data.frame(bind_rows(edges_0,edges_1,edges_2))
head(edges)
```

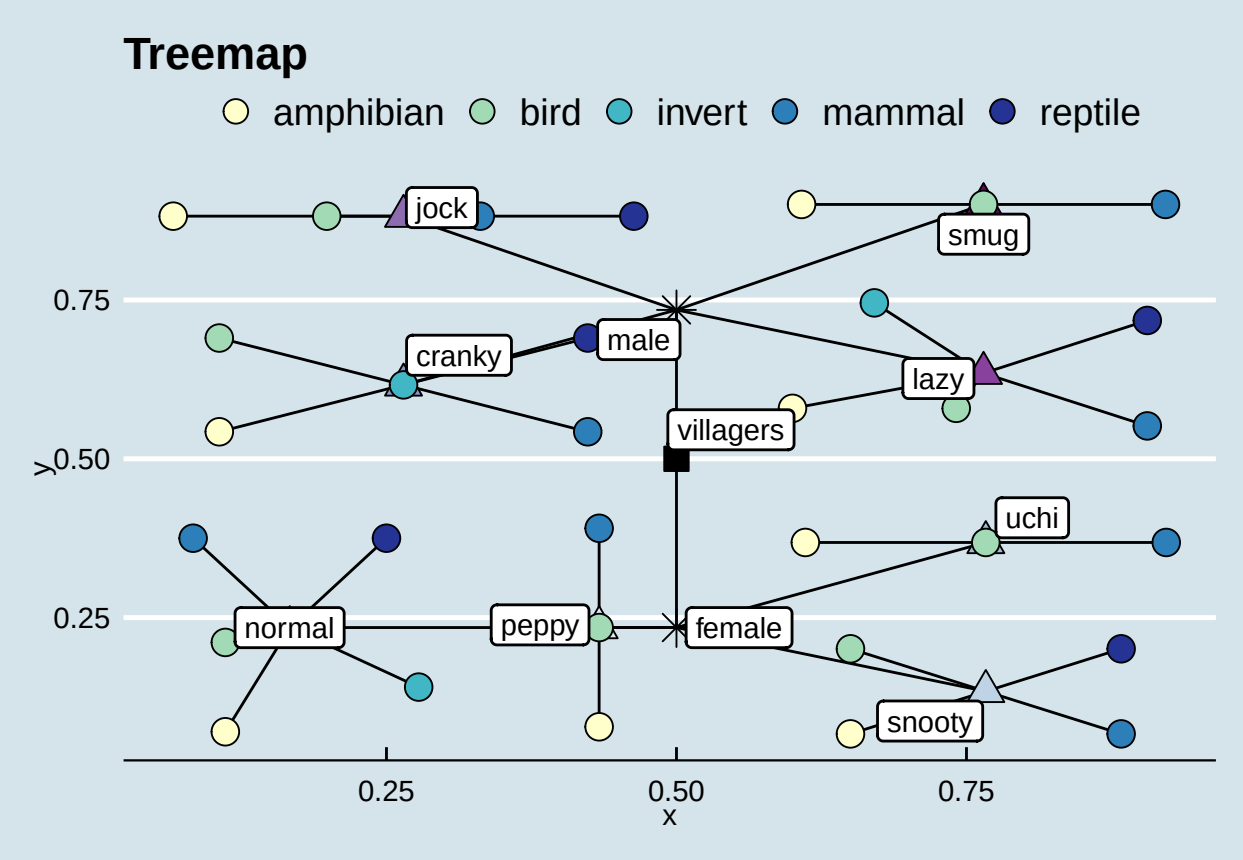
```
##                 to                    from
## 1        villagers         villagers.male
## 2        villagers       villagers.female
## 3   villagers.male  villagers.male.cranky
## 4 villagers.female villagers.female.peppy
## 5 villagers.female  villagers.female.uchi
```

```
## 6   villagers.male    villagers.male.lazy
```

```
gr<-graph_from_data_frame(edges,vertices=nodes)

#set up unique color pallet with different tones for different levels
cols<-c('black','gray','gray50',
        brewer.pal(length(unique(nodes$shortname[nodes$type=='personality'])),'BuPu'),
        brewer.pal(length(unique(nodes$shortname[nodes$type=='species'])),'YlGnBu'))
names(cols)<-unique(nodes$shortname)
```

```
p1<-ggraph(gr, 'treemap') +
    geom_edge_link(color='black') +
    geom_node_point(aes(fill = shortname,shape=type,size=4)) +
    scale_fill_manual(name='',breaks=c('amphibian','bird','invert','mammal','reptile'),values=cols) +
    scale_shape_manual(values=c(8,24,22,21)) +
    theme_economist() +
    geom_node_label(aes(label=label),repel = T,label.size=0.5,show.legend = NA) + guides(size=F,shape=F
    ggtitle('Treemap')
p1
```

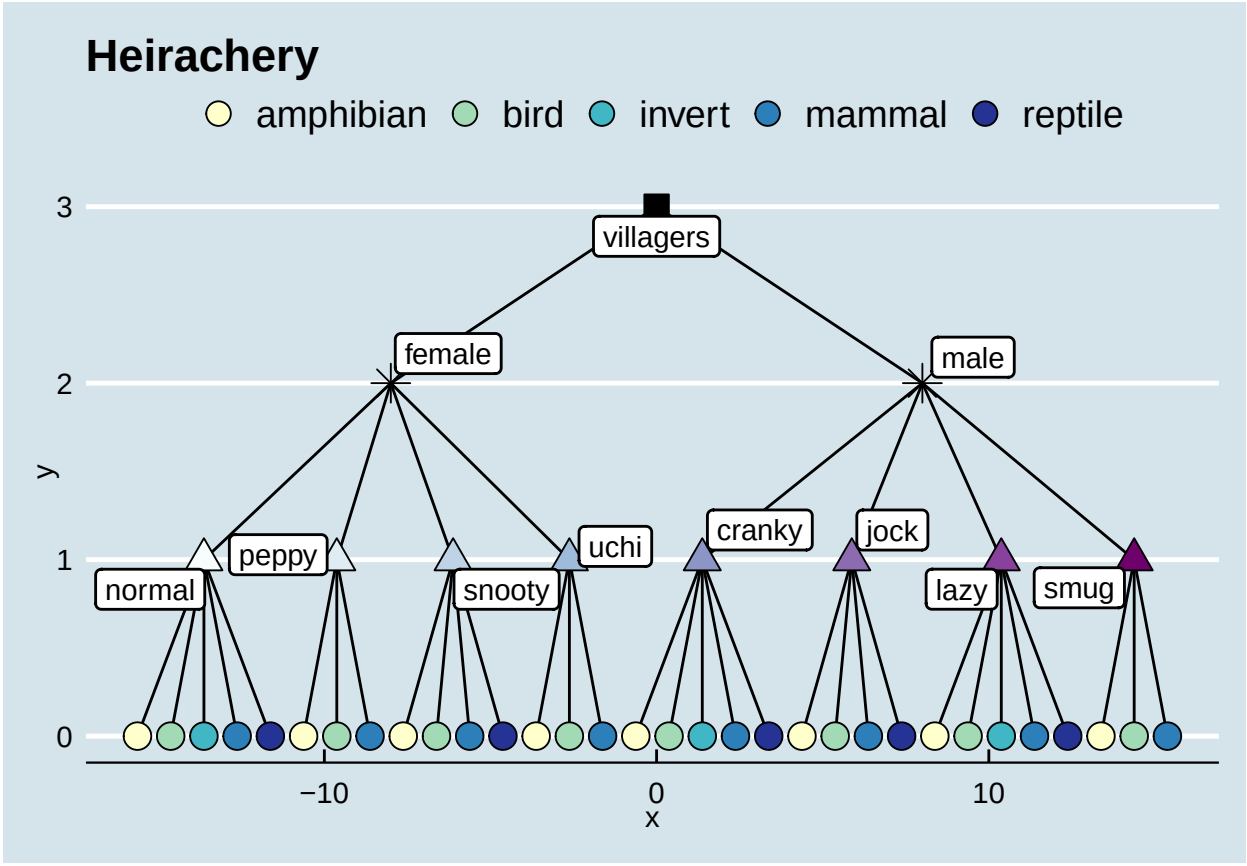



```
p2<-ggraph(gr) +
    geom_edge_link(color='black') +
    geom_node_point(aes(fill = shortname,shape=type,size=4)) +
    scale_fill_manual(name='',breaks=c('amphibian','bird','invert','mammal','reptile'),values=cols) +
    scale_shape_manual(values=c(8,24,22,21)) +
    theme_economist() +
    geom_node_label(aes(label=label),repel = T,label.size=0.5,show.legend = NA) + guides(size=F,shape=F
```

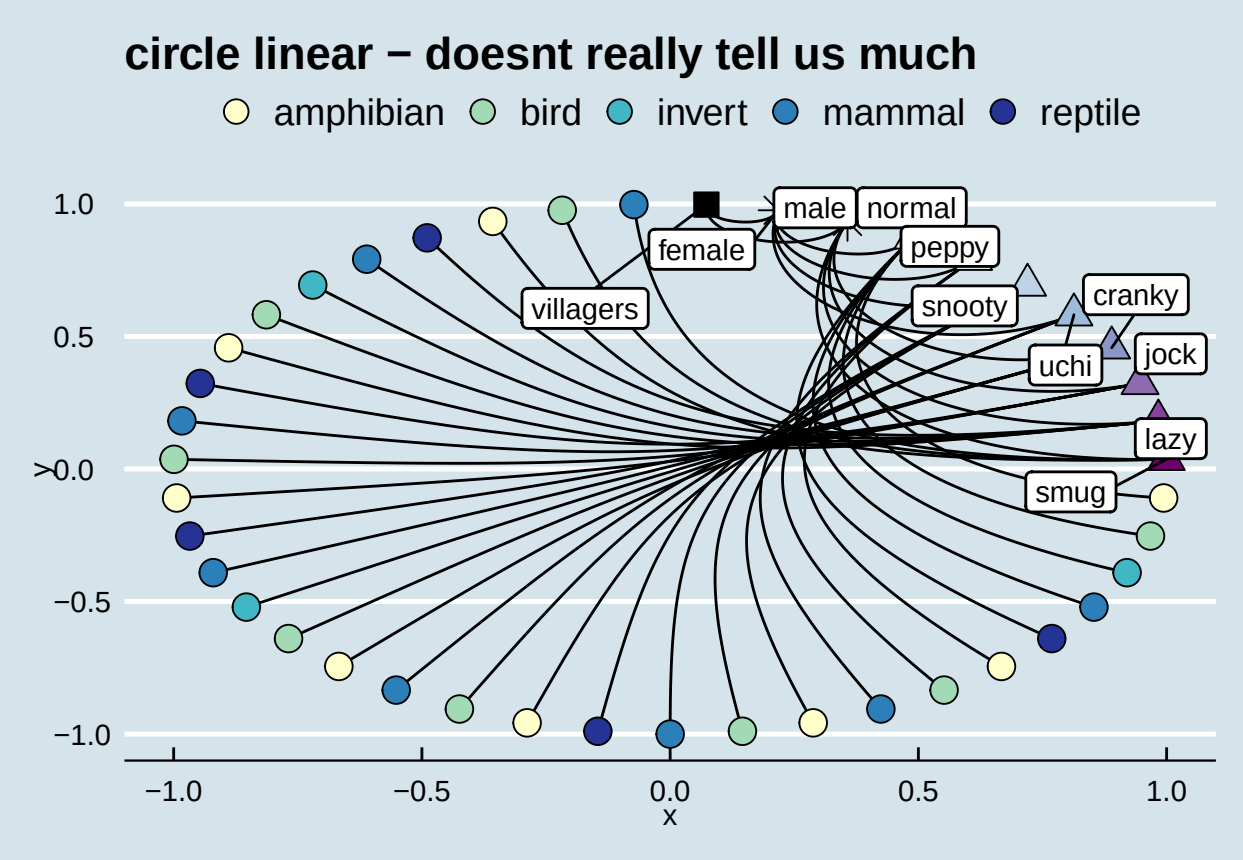
```
    ggtitle('Heirachery')
```

```
## Using `tree` as default layout
```

```
p2
```



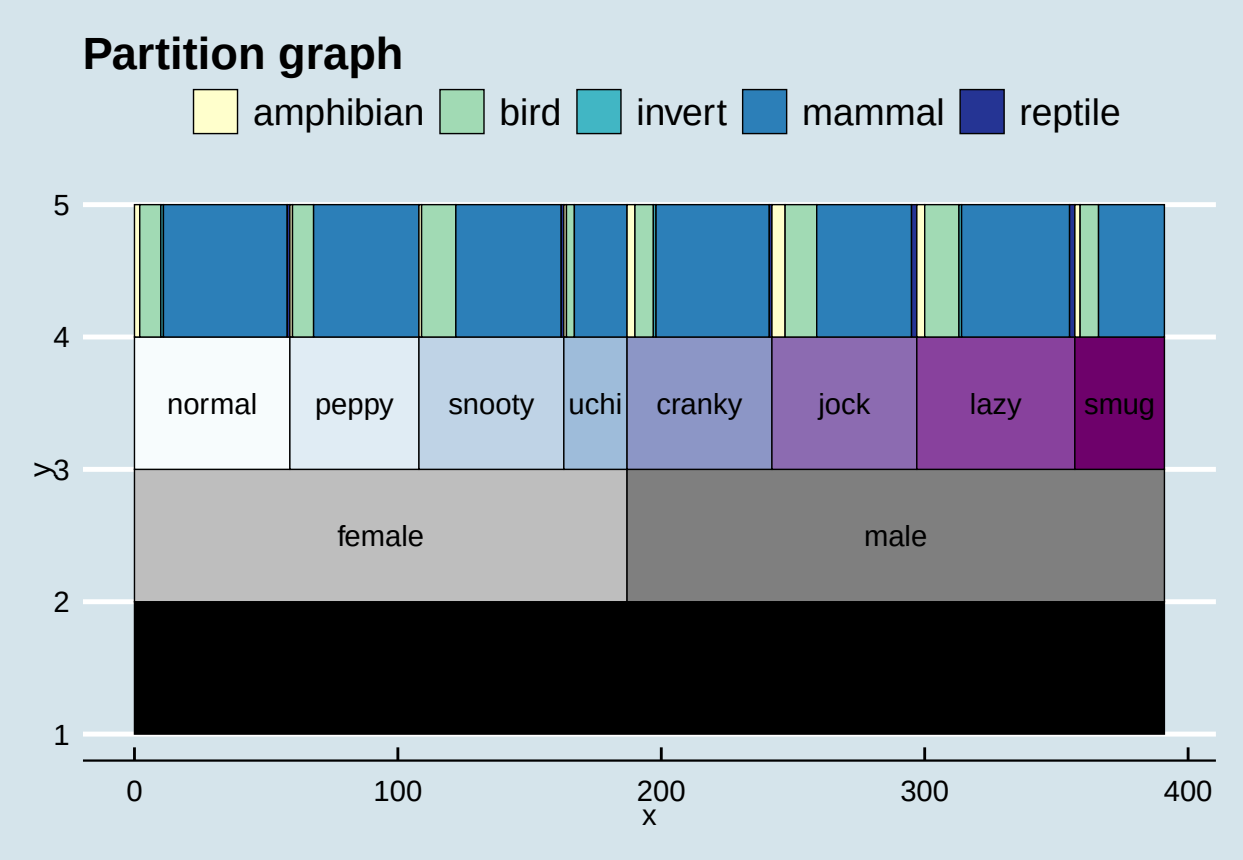

```
p3<-ggraph(gr,layout='linear',circular=T) +
    geom_edge_arc(color='black') +
    geom_node_point(aes(fill = shortname,shape=type,size=4)) +
    scale_fill_manual(name='',breaks=c('amphibian','bird','invert','mammal','reptile'),values=cols) +
    scale_shape_manual(values=c(8,24,22,21)) +
    theme_economist() +
    geom_node_label(aes(label=label),repel = T,label.size=0.5,show.legend = NA) + guides(size=F,shape=F
    ggtitle('circle linear - doesnt really tell us much')
p3
```




```
ggraph(gr, 'partition',weight=size) +
    geom_node_tile(aes(fill = shortname), size = 0.25) +
    scale_fill_manual(name='',breaks=c('amphibian','bird','invert','mammal','reptile'),values=cols) + g
  ggtitle('Partition graph')
```

```
## Non-leaf weights ignored
```

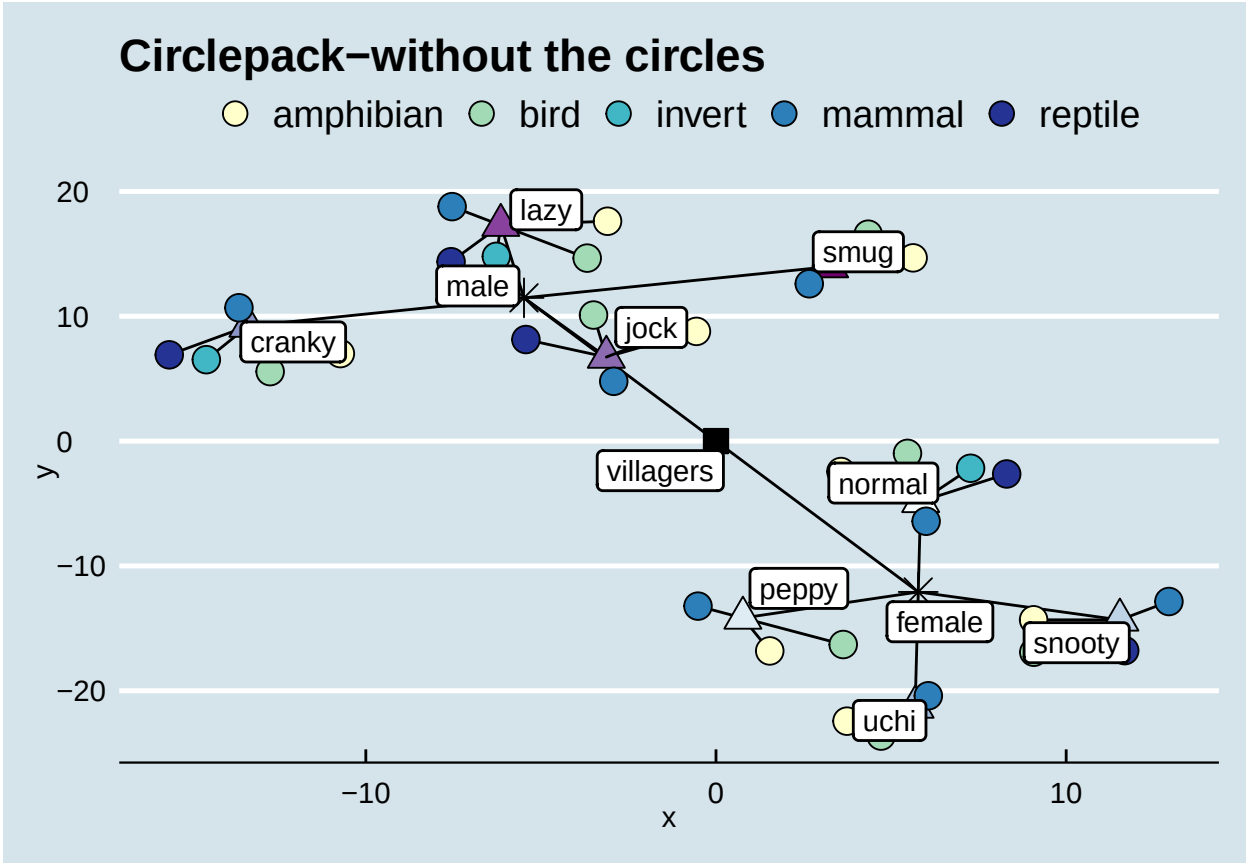
```
## Warning: Ignoring unknown parameters: label.size
```




```
set.seed(1)
p4<-ggraph(gr,'circlepack',weight=size) +
geom_edge_link(color='black') +
    geom_node_point(aes(fill = shortname,shape=type,size=4)) +
    scale_fill_manual(name='',breaks=c('amphibian','bird','invert','mammal','reptile'),values=cols) +
    scale_shape_manual(values=c(8,24,22,21)) +
    theme_economist() +
    geom_node_label(aes(label=label),repel = T,label.size=0.5,show.legend = NA) + guides(size=F,shape=F
    ggtitle('Circlepack-without the circles')
```

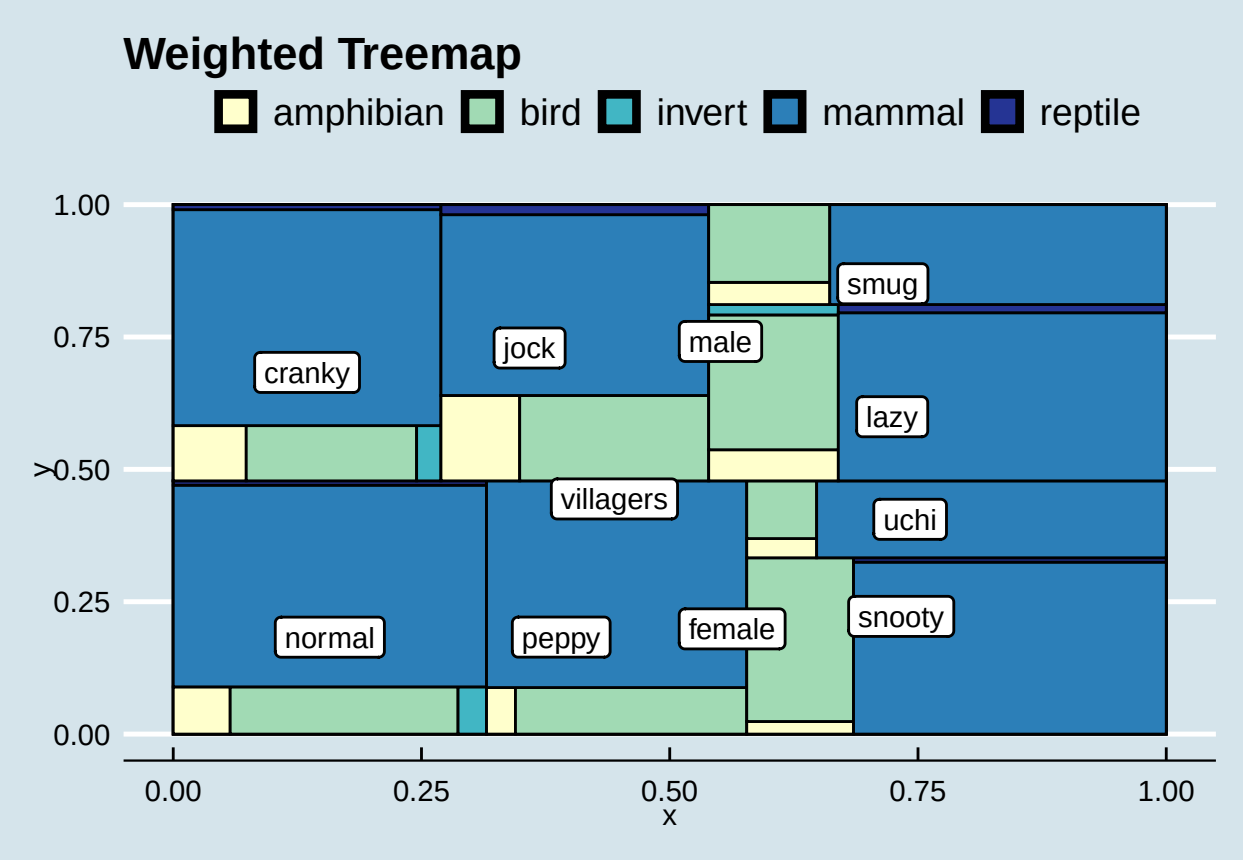
```
## Non-leaf weights ignored
```

```
p4
```

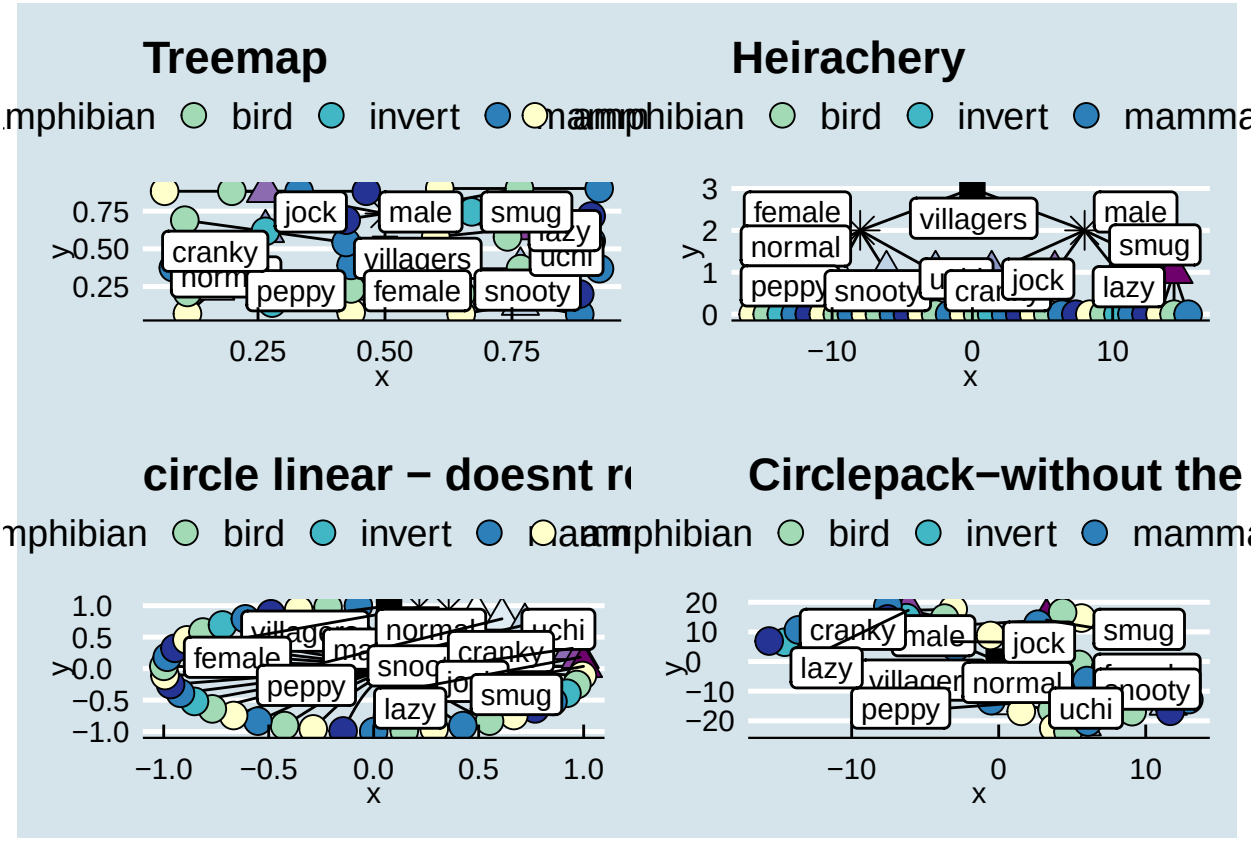



```
ggraph(gr,'treemap',weight=size) +
geom_node_tile(aes(fill=shortname)) +
    theme_economist() +
    geom_node_label(aes(label=label),repel = T,label.size=0.5,show.legend = NA) +
  scale_fill_manual(name='',breaks=c('amphibian','bird','invert','mammal','reptile'),values=cols) +
  guides(size=F,shape=F,fill=guide_legend(override.aes = list(shape=21,size=4))) +
    ggtitle('Weighted Treemap')
```

```
## Non-leaf weights ignored
```




```
(p1 + p2) / (p3 + p4) + ggsave('animal_crossings.jpeg',width=10,height=10)
```


Treemap
mphibian
bird
invert
jock
male
smug
cranky
villagers
peppy
female
snooty
0.75
0.50
0.25
y
0.25
0.50
0.75
x
Heirachery
amphibian
bird
invert
mamma
female
villagers
male
normal
smug
peppy
snooty
jock
lazy
3
2
1
0
y
−10
0
10
x
circle linear − doesnt r
mphibian
bird
invert
female
peppy
lazy
smug
1.0
0.5
0.0
−0.5
−1.0
y
−1.0
−0.5
0.0
0.5
1.0
x
Circlepack−without the
amphibian
bird
invert
mamma
cranky
male
jock
smug
lazy
normal
peppy
uchi
20
10
0
−10
−20
y
−10
0
10
x