

jesse_linkage_analysis

Jesse Cahill

12/1/2018

Appliances

Tidying

```
# appliance data

appliances <- read_tun_data(here("data/raw",
                                "Households by possession of electrical household equipments 11_18_2018.csv"),
  mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names
  filter(Region != "Tunisia")

# household / population data
house_pop <- read_csv(here("data/intermediate",
                           "master_file.csv")) %>%

  rename(Region = governorate)

## Parsed with column specification:
## cols(
##   governorate = col_character(),
##   over_15_population = col_integer(),
##   population = col_integer(),
##   housing = col_integer(),
##   households = col_integer()
## )

appliances <- inner_join(select(house_pop, Region, households), appliances)

## Joining, by = "Region"

# tidy
appliances <- appliances %>%
  gather(contains("_"), key="key", value="value") %>%
  separate(key, into=c("Appliance", "Measure"), sep="_") %>%
  spread(Appliance, value)

# we only want total
appliances_total <- appliances %>%
  filter(Measure == "Total") %>%
  select(-Measure)

# fix the names
ap_names = c("Region", "households", "AC", "heat", "dishwasher", "stove", "fridge", "washing_machine")
names(appliances_total) <- ap_names

# cleaning up
appliances_total <- appliances_total %>%
```

```

mutate_at(.vars = vars(-Region), funs(as.numeric)) %>%
mutate(Region = str_replace(Region, "Governorate of ", ""))

#plots

appliances_tidy <- appliances_total %>%
  gather(-Region, -households, key="appliances_type", value="num_appliances") %>%
  mutate(appliances_household = num_appliances / households)

g_stacked <- ggplot(appliances_tidy, aes(y = appliances_household,
                                         x = fct_reorder2(Region, appliances_type, -appliances_household),
                                         fill = appliances_type)) +
  coord_flip() +
  geom_bar(stat="identity") + ylab("") +
  ggtitle("appliances Resources per Household")

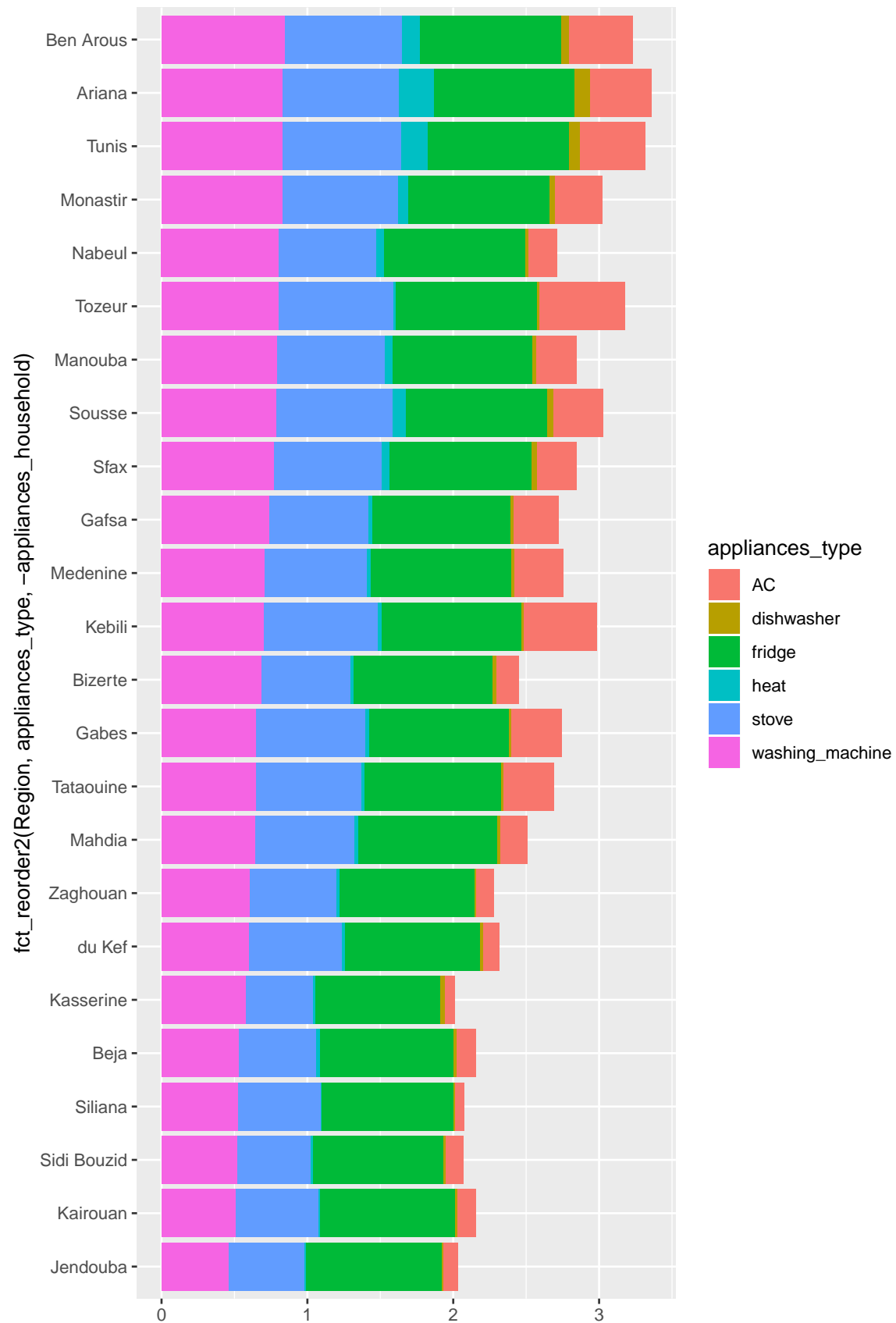
g_dot <- ggplot(appliances_tidy, aes(x = appliances_household,
                                     y = fct_reorder2(Region, appliances_type, -appliances_household),
                                     color = appliances_type)) +
  geom_point() + ylab("") +
  ggtitle("appliances Resources per Household")

g_dot_facet <- ggplot(appliances_tidy, aes(x = appliances_household,
                                           y = reorder(Region, appliances_household))) +
  geom_point() + ylab("") +
  ggtitle("appliances Resources per Household")

g_stacked

```

appliances Resources per Household



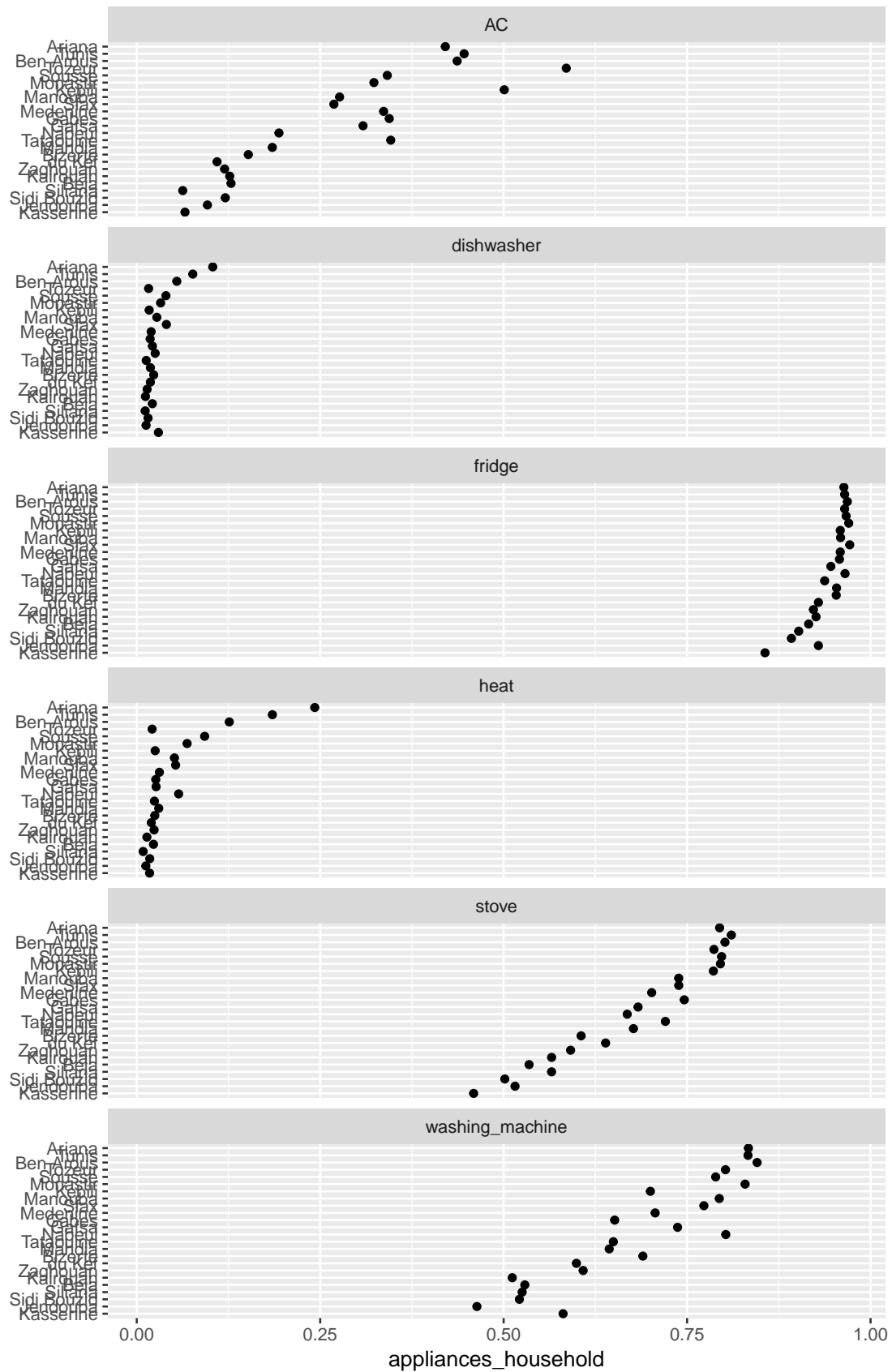
g_dot

appliances Resources per Household



```
g_dot_facet + facet_wrap(~appliances_type, ncol=1)
```

appliances Resources per Household



Internet Connections

```
internet <- read_tun_data(here("data/raw",
                             "Households by possession of ICTs 11_18_2018 02_51_58.xlsx")) %>%
  mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names
  filter(Region != "Tunisia") %>%
  select(-contains("households_"))

internet <- inner_join(select(house_pop, Region, households), internet)

## Joining, by = "Region"

# tidy
internet <- internet %>%
  gather(contains("_"), key="key", value="value") %>%
  separate(key, into=c("Appliance", "Measure"), sep="_") %>%
  spread(Appliance, value)

# we only want total
internet_total <- internet %>%
  filter(Measure == "Total") %>%
  select(-Measure)

#fix the names
internet_names = c("Region", "households", "computer", "landline", "internet", "mobile")
names(internet_total) <- internet_names

# cleaning up
internet_total <- internet_total %>%
  mutate_at(.vars = vars(-Region), funs(as.numeric)) %>%
  mutate(Region = str_replace(Region, "Governorate of ", ""))

#graph it

internet_tidy <- internet_total %>%
  gather(-Region, -households, key="ict_type", value="num_ict") %>%
  mutate(ict_household = num_ict / households)

g_stacked <- ggplot(internet_tidy, aes(y = ict_household,
                                       x = fct_reorder2(Region, ict_type, -ict_household),
                                       fill = ict_type)) +
  coord_flip() +
  geom_bar(stat="identity") + ylab("") +
  ggtitle("ICT Resources per Household")

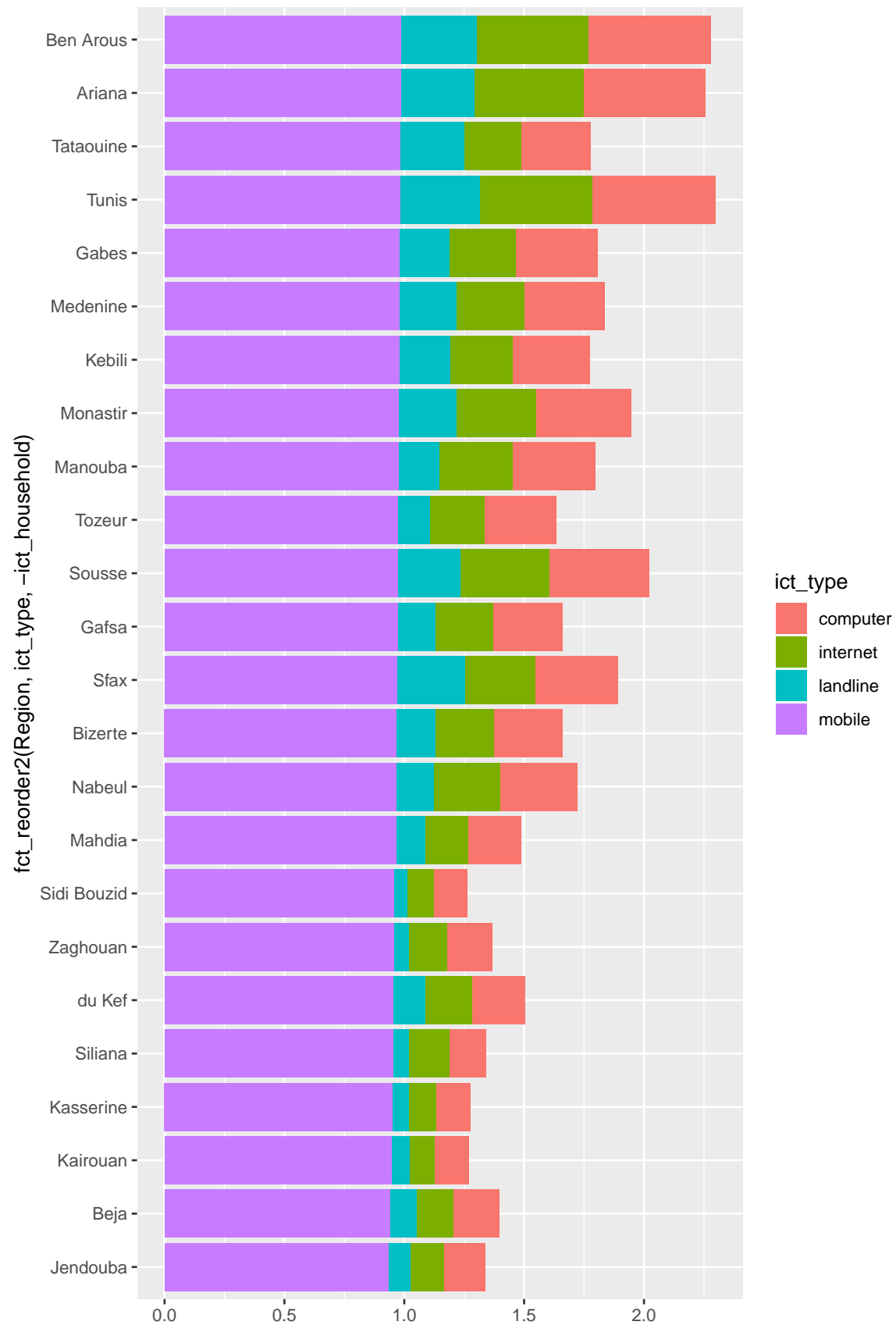
g_dot <- ggplot(internet_tidy, aes(x = ict_household,
                                   y = fct_reorder2(Region, ict_type, -ict_household),
                                   color = ict_type)) +
  geom_point() + ylab("") +
  ggtitle("ICT Resources per Household")

g_dot_facet <- ggplot(internet_tidy, aes(x = ict_household,
                                          y = reorder(Region, ict_household))) +
  geom_point() + ylab("") +
```



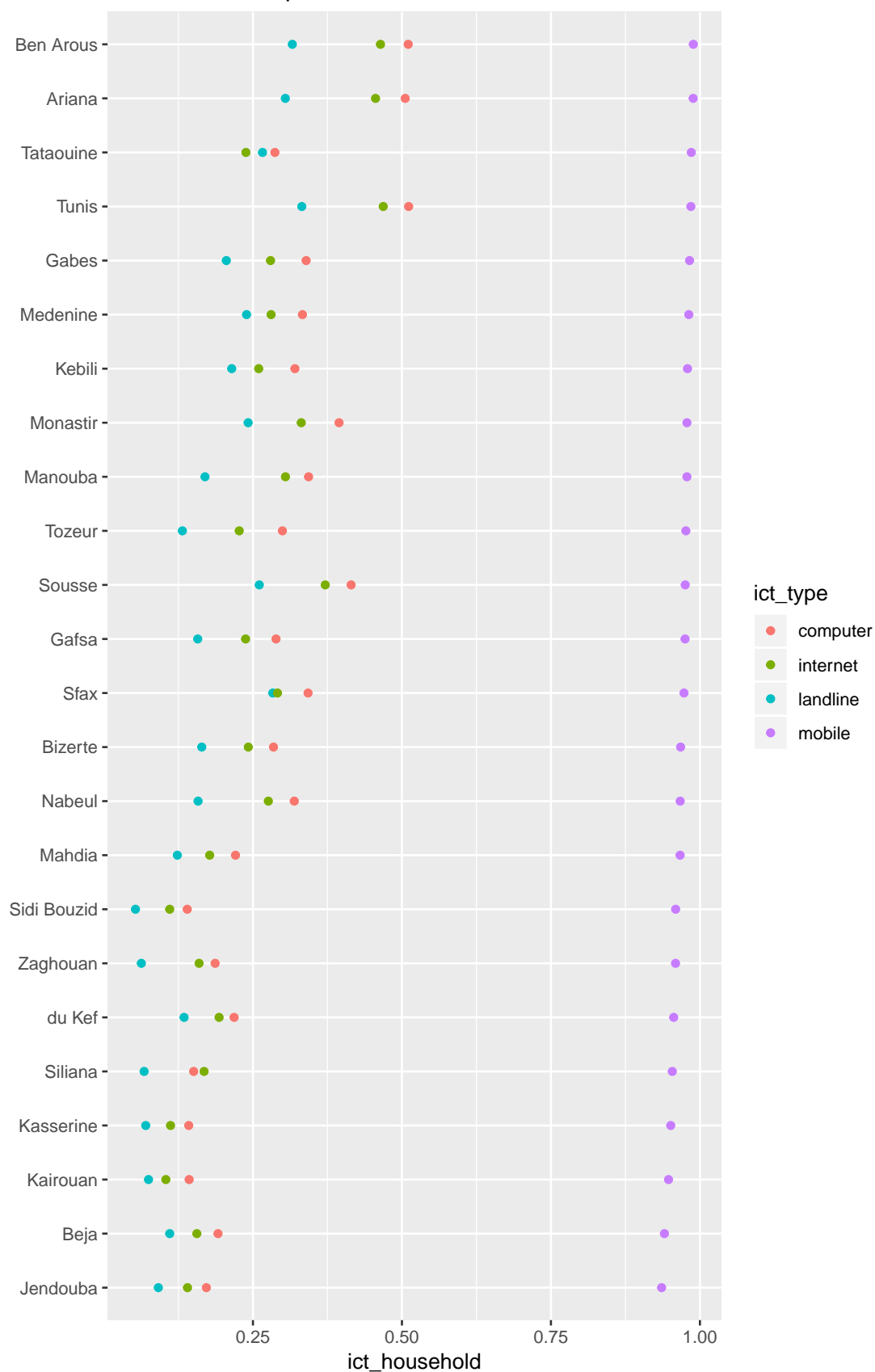
```
ggtitle("internet Resources per Household")  
g_stacked
```

ICT Resources per Household



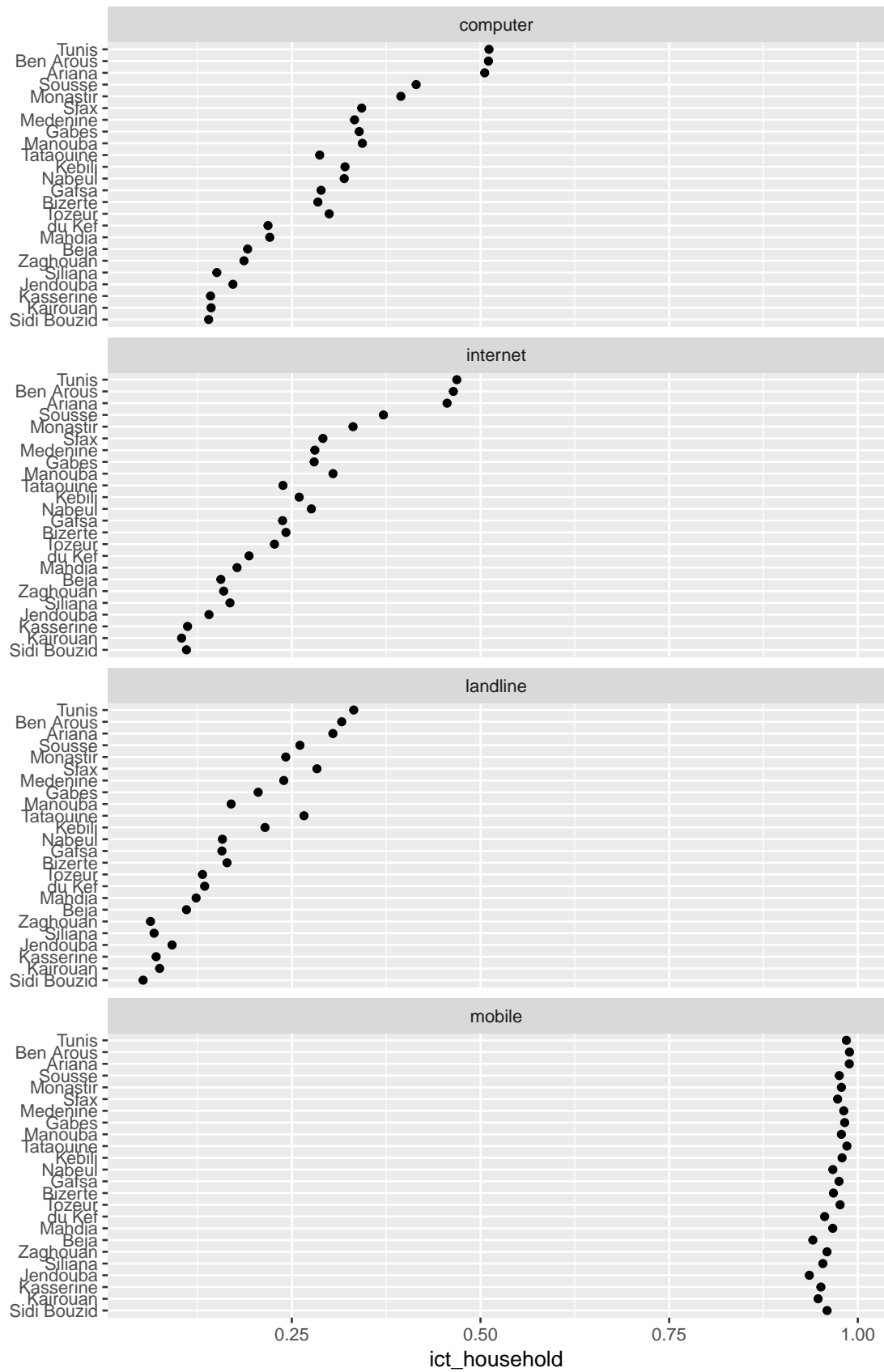
g_dot

ICT Resources per Household



```
g_dot_facet + facet_wrap(~ict_type, ncol=1)
```

internet Resources per Household

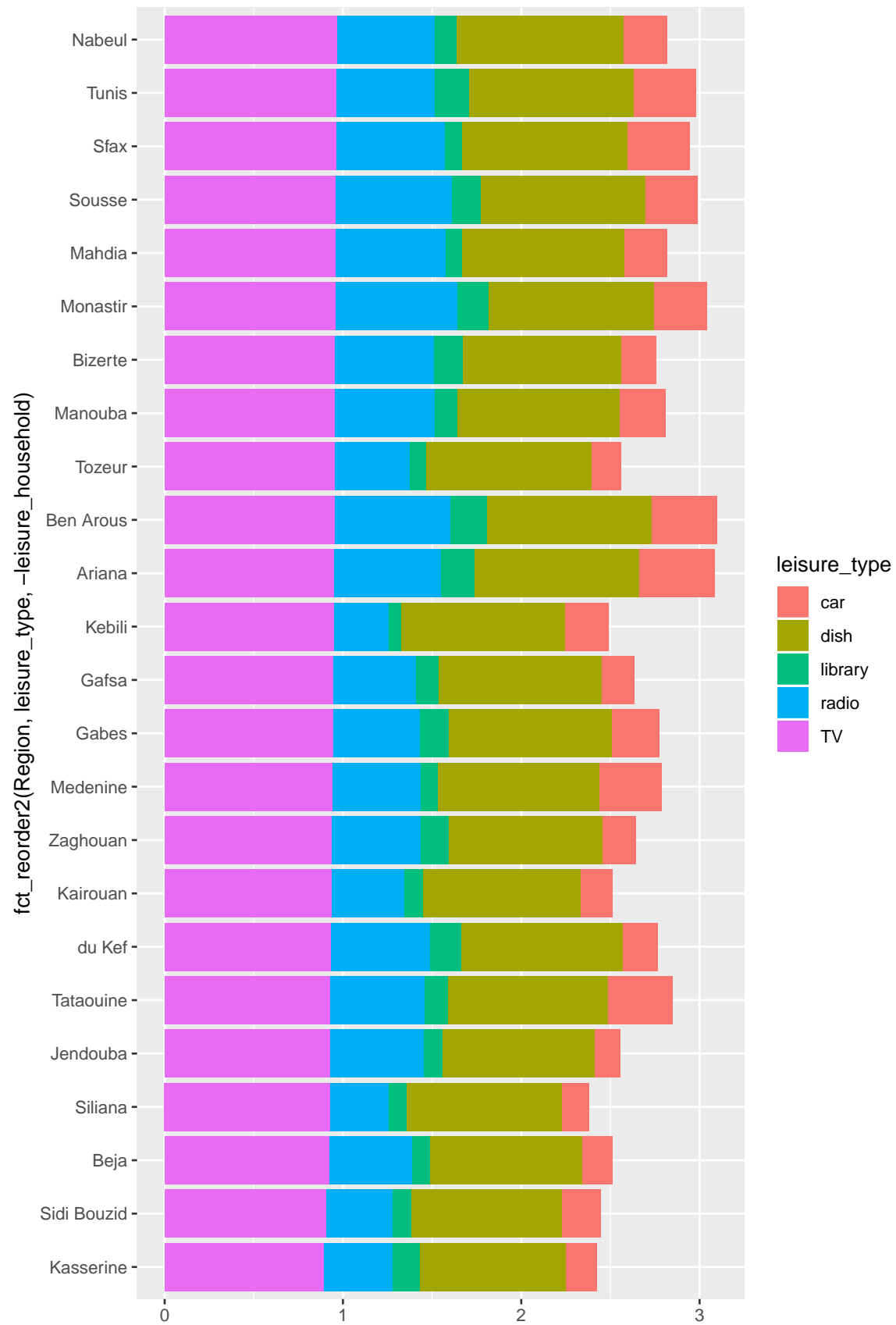


Leisure Resources

```
leisure <- read_tun_data(here("data/raw",  
                             "Households by possession of leisure resources 11_18_2018 02_51_43.xls"),  
  mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names  
  filter(Region != "Tunisia") %>%  
  select(-contains("households_"))  
  
leisure <- inner_join(select(house_pop, Region, households), leisure)  
  
## Joining, by = "Region"  
  
# tidy  
leisure <- leisure %>%  
  gather(contains("_"), key="key", value="value") %>%  
  separate(key, into=c("Appliance", "Measure"), sep="_") %>%  
  spread(Appliance, value)  
  
# we only want total  
leisure_total <- leisure %>%  
  filter(Measure == "Total") %>%  
  select(-Measure)  
  
# fix the names  
leisure_names = c("Region", "households", "car", "library", "dish", "radio", "TV")  
names(leisure_total) <- leisure_names  
  
# cleaning up  
leisure_total <- leisure_total %>%  
  mutate_at(.vars = vars(-Region), funs(as.numeric)) %>%  
  mutate(Region = str_replace(Region, "Governorate of ", ""))  
  
leisure_tidy <- leisure_total %>%  
  gather(-Region, -households, key="leisure_type", value="num_leisure") %>%  
  mutate(leisure_household = num_leisure / households)  
  
g_stacked <- ggplot(leisure_tidy, aes(y = leisure_household,  
                                     x = fct_reorder2(Region, leisure_type, -leisure_household),  
                                     fill = leisure_type)) +  
  coord_flip() +  
  geom_bar(stat="identity") + ylab("") +  
  ggtitle("Leisure Resources per Household")  
  
g_dot <- ggplot(leisure_tidy, aes(x = leisure_household,  
                                 y = fct_reorder2(Region, leisure_type, -leisure_household),  
                                 color = leisure_type)) +  
  geom_point() + ylab("") +  
  ggtitle("Leisure Resources per Household")  
  
g_dot_facet <- ggplot(leisure_tidy, aes(x = leisure_household,  
                                       y = reorder(Region, leisure_household))) +  
  geom_point() + ylab("") +  
  ggtitle("Leisure Resources per Household")
```

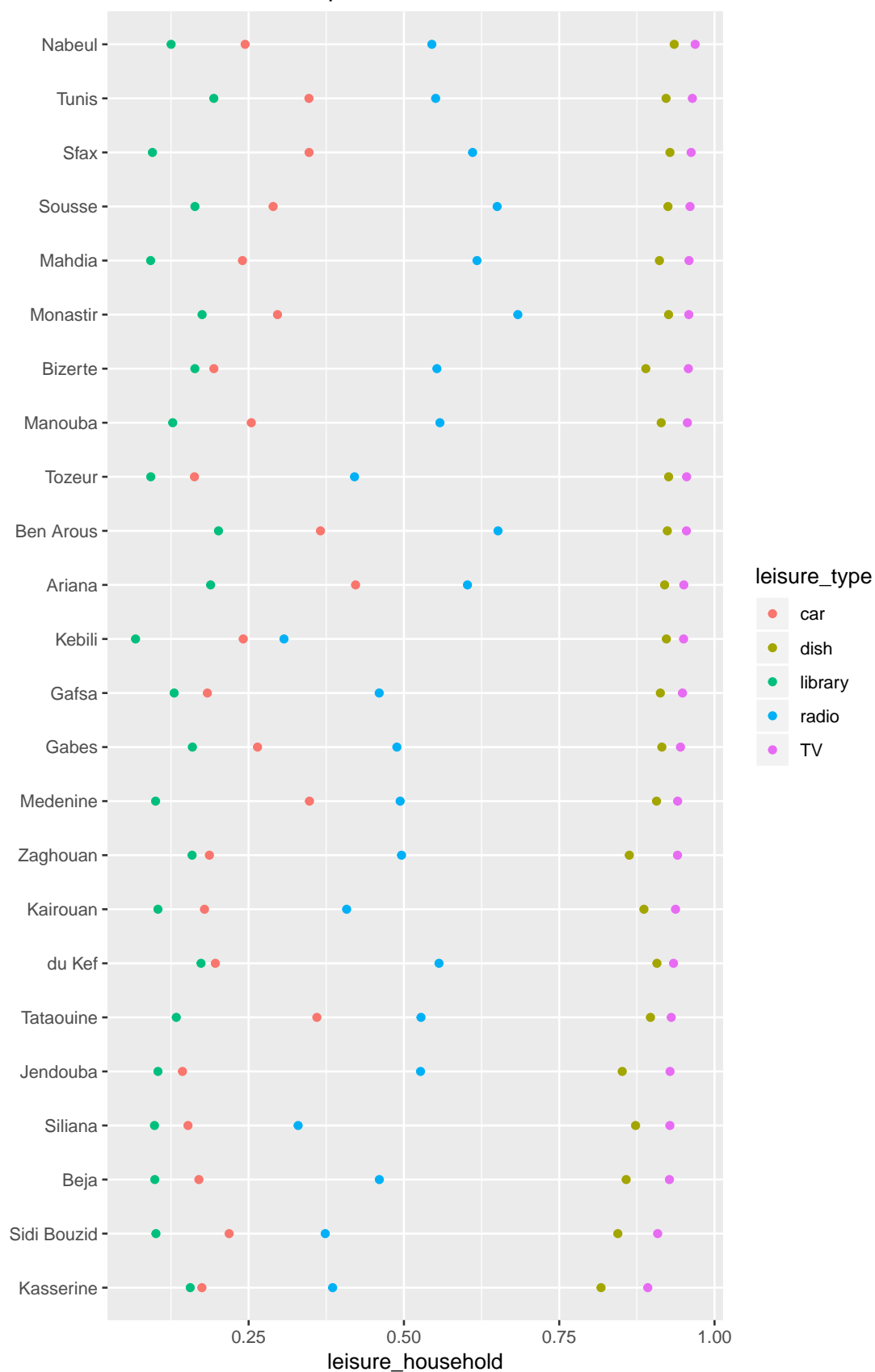
g_stacked

Leisure Resources per Household



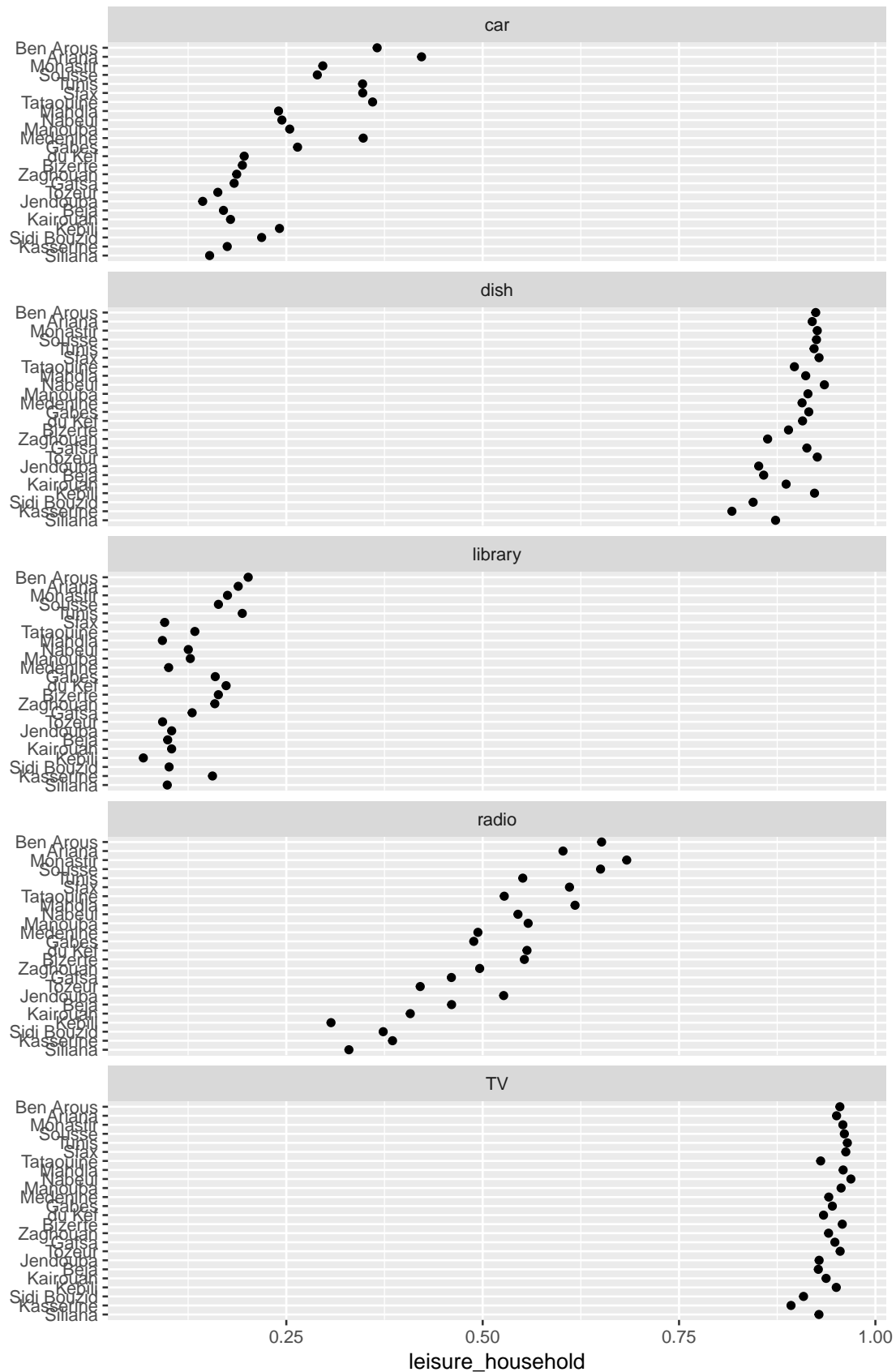
g_dot

Leisure Resources per Household



```
g_dot_facet + facet_wrap(~leisure_type, ncol=1)
```

Leisure Resources per Household



Distribution of Occupation by activity sector

Parallel Coordinate Plot to Show how Activity Sectors covary

```
occupation <- read_tun_data(here("data/raw",  
                                "Distribution of the occupied population aged 15 years and over by the  
                                mutate(Region = str_replace(Region, "Governorate of ", "")) %>% # Standardize governorate names  
                                filter(Region != "Tunisia")  
  
occupation <- occupation %>%  
  mutate_at(vars(contains("_")), function(x) as.numeric(x) / house_pop$over_15_population) %>%  
  select(-ends_with("Total"))  
  
occupation <- occupation %>%  
  gather(contains("_"), key="key", value="value") %>%  
  separate(key, into=c("sector", "gender"), sep="_") %>%  
  spread(sector, value)  
  
library(GGally)  
  
##  
## Attaching package: 'GGally'  
  
## The following object is masked from 'package:dplyr':  
##  
##      nasa  
  
# add y label - percent of total pop over 15 employed  
occ_names <- c("Region", "gender", "Agr./Fishing", "Pub. Works", "Education/Health/Admin", "Manufacturing",  
              "Other Serv.", "Trade", "Transport", "Undeclares")  
names(occupation) <- occ_names  
ggparcoord(occupation, columns = 3:11, alphaLines = .5,  
           scale = "uniminmax", splineFactor = 10, groupColumn = 2) +  
  geom_vline(xintercept = 2:9, color = "lightblue")
```

