SSPs Human Capitals - National

James Millington

Libraries

```
library(tidyr)
library(dplyr)
library(readxl)
library(ggplot2)
library(ggrepel)
library(ggiraph)
```

Load Data

```
load_ssp <- function(filepath,varlab){
  #for each SSP data sheet
  for(n in 1:5){
    ssp <- paste0("ssp", n)
    sheet <- read_excel(filepath, sheet = ssp, na='<Null>')

    #pivot longer, add ssp id column
    lsheet <- sheet %>%
      pivot_longer(
        cols = !OID:unit,
        names_to='year',
        values_to='value'
      ) %>%
      mutate(ssp=n)

    #calculate ranks of values
    lranks <- lsheet %>%
      group_by(year) %>%
```

```
      mutate(unit='rank',
             #value=min_rank(desc(value)))   #high rank number is low value
             value=min_rank(value))   #high rank number is high value

    #combine into single dataframe
    ldat <- rbind(lsheet,lranks)
    if(n > 1){
      alldat <- rbind(alldat,ldat)
    } else { alldat <- ldat}
  }
  alldat <- mutate(alldat,variable=varlab)
  return(alldat)
}
```

```
edu <- load_ssp("data/edu.xlsx", "edu")
ma <- load_ssp("data/ma.xlsx", "ma")
gdp <- load_ssp("data/gdp.xlsx", "gdp")
health <- load_ssp("data/health.xlsx", "health")
gini <- load_ssp("data/gini.xlsx", "gini")
wap <- load_ssp("data/wap.xlsx", "wap")
tec <- load_ssp("data/tec.xlsx", "tec")

countries <-edu %>%
  select(OID, GID_0) %>%
  distinct()
```

Calculations

```
#root mean square error (using median)
rmse_med <- function(x) {
  med = median(x, na.rm=TRUE)
  return(sum(sqrt((x-med)^2)))
}

#range function
rangesr <- function(x) {
  maxv = max(x, na.rm=TRUE)
  minv = min(x, na.rm=TRUE)
  return(diff(c(minv,maxv)))
```

```
}

out_all = countries
for(nm in list(edu,ma, gdp, health, gini, wap, tec)){
  rmsemed_nm = paste0(nm$variable[1],"_rmsemed")
  medmed_nm = paste0(nm$variable[1],"_medmed")
  summed_nm = paste0(nm$variable[1],"_summed")
  sumrng_nm = paste0(nm$variable[1],"_sumrng")
  summax_nm = paste0(nm$variable[1],"_summax")
  summin_nm = paste0(nm$variable[1],"_summin")

  out_all <- nm %>%
  #edu %>%
    filter(unit=='rank') %>%
    filter(!is.na(value)) %>%
    group_by(GID_0, ssp) %>%
    #calc median rank across years, for each ssp, for each country
    summarise(medrank = median(value, na.rm=TRUE),
              maxrank = max(value, na.rm=TRUE),
              minrank = min(value, na.rm=TRUE),
              rangerank = rangesr(value),
              ) %>%
    group_by(GID_0) %>%
    summarise(!!rmsemed_nm := rmse_med(medrank), #calc rmse (median) for ssp medians
              !!medmed_nm := median(medrank),    #calc median of ssp medians
              !!summed_nm := sum(medrank),       #calc sum of ssp medians
              !!sumrng_nm := sum(rangerank),     #calc sum of ssp ranges
              !!summax_nm := sum(maxrank),       #calc sum of ssp maxs
              !!summin_nm := sum(minrank),       #calc sum of ssp mins
              ) %>%

    left_join(out_all, ., by='GID_0')
}


out_2100 = countries
for(nm in list(edu,ma, gdp, health, gini, wap, tec)){
  rmsemed_nm = paste0(nm$variable[1],"_rmsemed")  #calc rmse (median) for ssp medians
  med_nm = paste0(nm$variable[1],"_med")
  rng_nm = paste0(nm$variable[1],"_rng")
  max_nm = paste0(nm$variable[1],"_max")
```

```
  min_nm = paste0(nm$variable[1],"_min")
  sum_nm = paste0(nm$variable[1],"_sum")

  out_2100 <- nm %>%
  #edu %>%
    filter(unit=='rank', year==2100)  %>%
    filter(!is.na(value)) %>%
    group_by(GID_0) %>%
    summarise(!!rmsemed_nm := rmse_med(value),       ##calc rmse (median) for ssp ranks
              !!med_nm := median(value, na.rm=TRUE), #calc median of ssp 2100 ranks
              !!sum_nm := sum(value),                #calc sum of ssp 2100 ranks
              !!rng_nm := rangesr(value),            #calc range of ssp 2100 ranks
              !!max_nm := max(value, na.rm=TRUE),    #calc sum of ssp 2100 rank maxs
              !!min_nm := min(value, na.rm=TRUE),    #calc sum of ssp 2100 rank mins
              ) %>%

    left_join(out_2100, ., by='GID_0')
}
```

Plots

```
#setup template https://stackoverflow.com/a/16727357

pall <-
  list(
    geom_point(alpha=0.75),
    ggtitle("SSPs over all Years"),
    scale_colour_gradient(low="darkblue",high="darkgrey"),
    xlim(-5, NA),
    ylim(-5, NA),
    geom_abline(intercept = 0, slope = 1,
                linewidth = 0.35,colour='red', linetype='dashed'),
    geom_label_repel(aes(label = GID_0),
                     alpha=0.5,
                     max.overlaps=15,
                     box.padding   = 0.35,
                     point.padding = 0.5,
                     segment.color = 'grey50',
                     show.legend = FALSE,
                     color='black'),
```

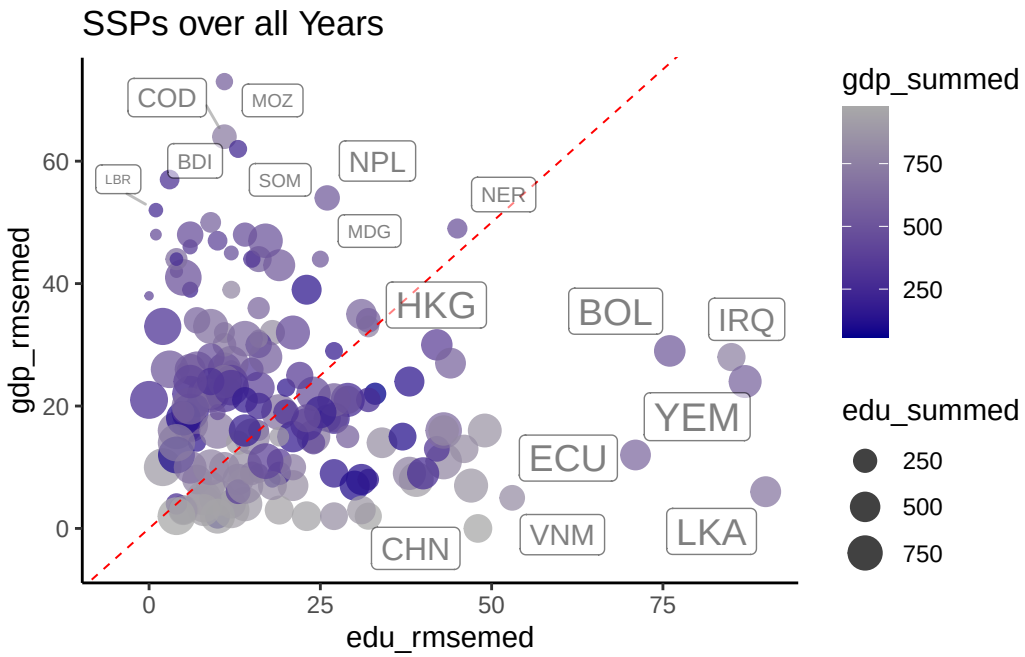
```
    theme_classic()
  )
```

```
#setup template https://stackoverflow.com/a/16727357

p2100 <-
  list(
    geom_point(alpha=0.75),
    ggtitle("SSPs in 2100"),
    scale_colour_gradient(low="darkblue",high="darkgrey"),
    xlim(-5, NA),
    ylim(-5, NA),
    geom_abline(intercept = 0, slope = 1,
                linewidth = 0.35,colour='red', linetype='dashed'),
    geom_label_repel(aes(label = GID_0),
                     alpha=0.5,
                     max.overlaps=15,
                     box.padding   = 0.35,
                     point.padding = 0.5,
                     segment.color = 'grey50',
                     show.legend = FALSE,
                     color='black'),
    theme_classic()
  )
```

Education

All Years

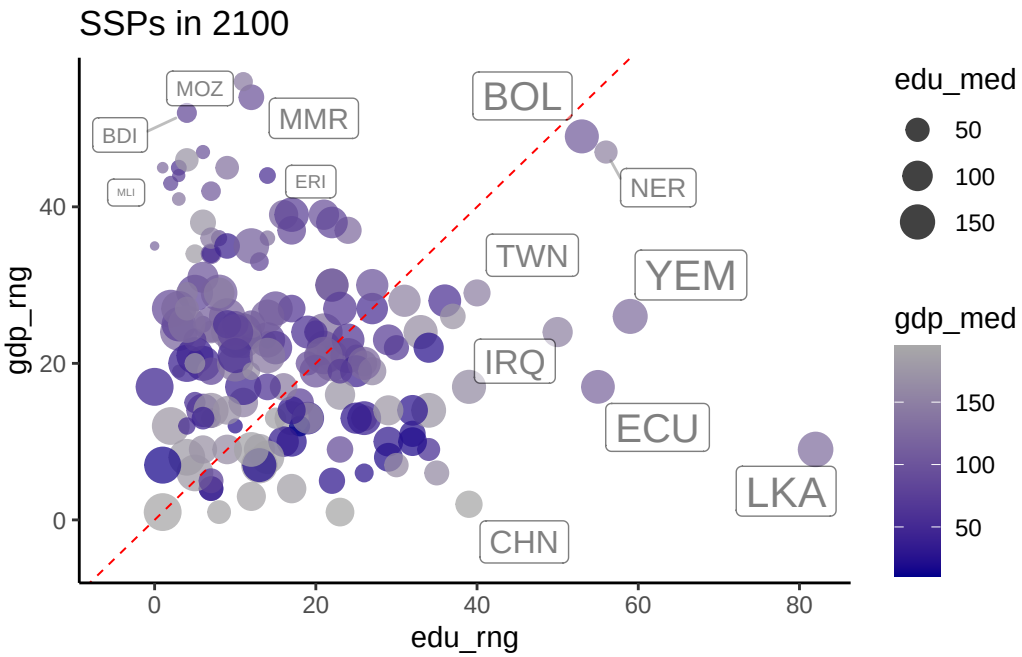
```
out_all %>%
  drop_na() %>%
  ggplot(aes(x=edu_rmsemed, y=gdp_rmsemed, size=edu_summed, colour=gdp_summed)) +
  pall
```



More grey is higher overall GDP ranking (e.g. USA is in bottom left)

2100

```
out_2100 %>%
  drop_na() %>%
  ggplot(aes(x=edu_rng, y=gdp_rng, size=edu_med, colour=gdp_med)) +
  p2100
```



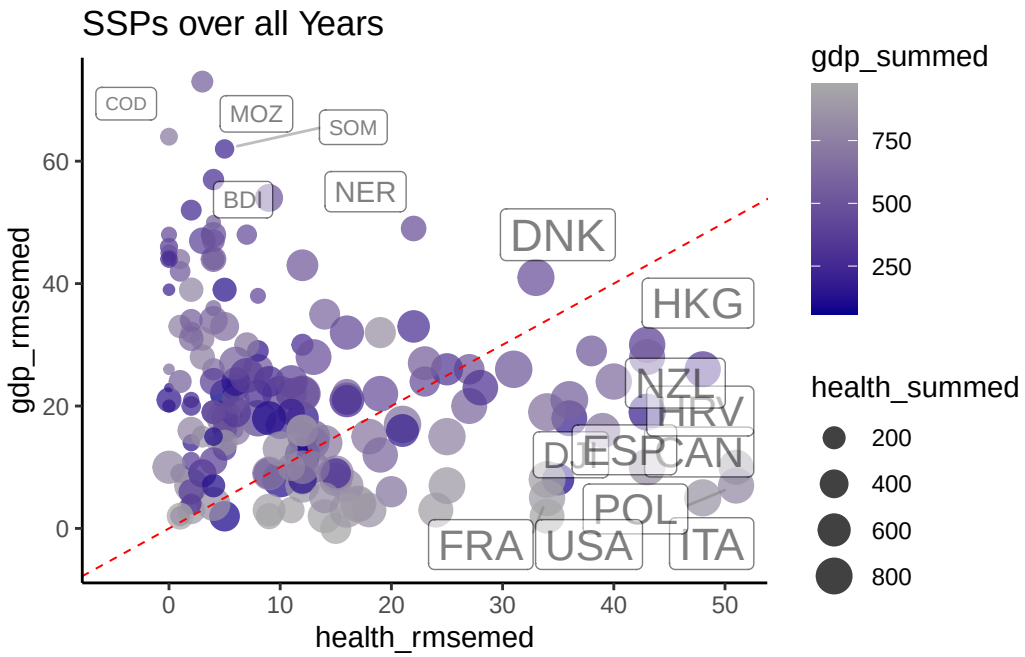
More grey is higher overall GDP ranking (e.g. USA is in bottom left)

Interpretation

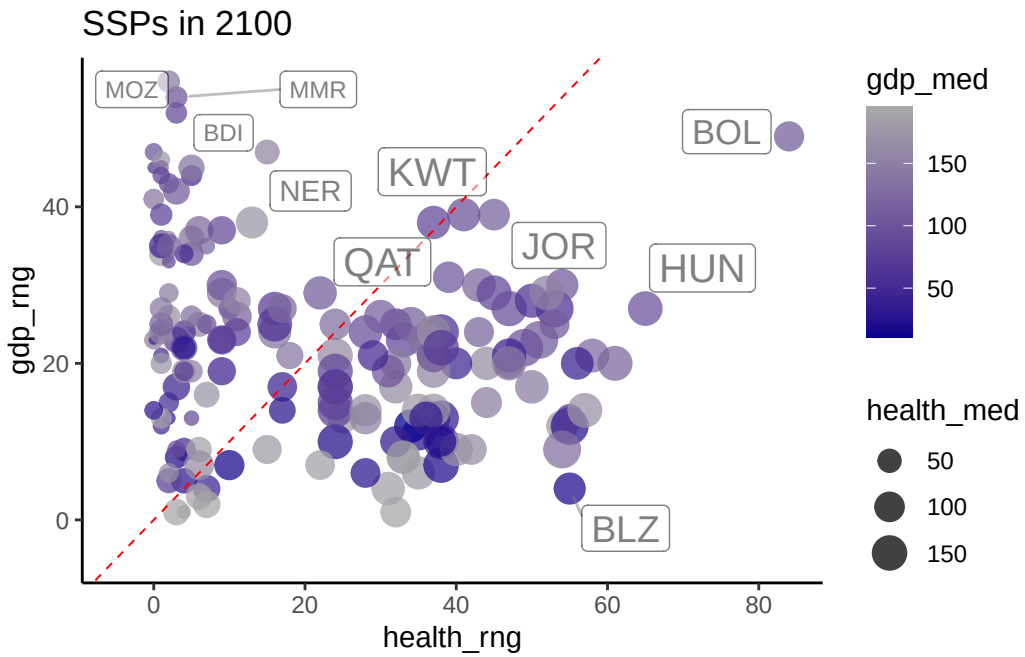
- China has low GDP range (always ranked well in 2100), but relatively variable Educational rank
- USA always ranked well on both indicators
- Sri Lanka (LKA) has a relatively consistent GDP ranking (quite high), but highly variable education ranking
- Mozambique, Burundi, Myanmar have variable GDP ranking (intermediate), but relatively consistent (poor) education ranking

Health

```
out_all %>%
  drop_na() %>%
  ggplot(aes(x=health_rmsemed, y=gdp_rmsemed, size=health_summed, colour=gdp_summed)) +
  pall
```

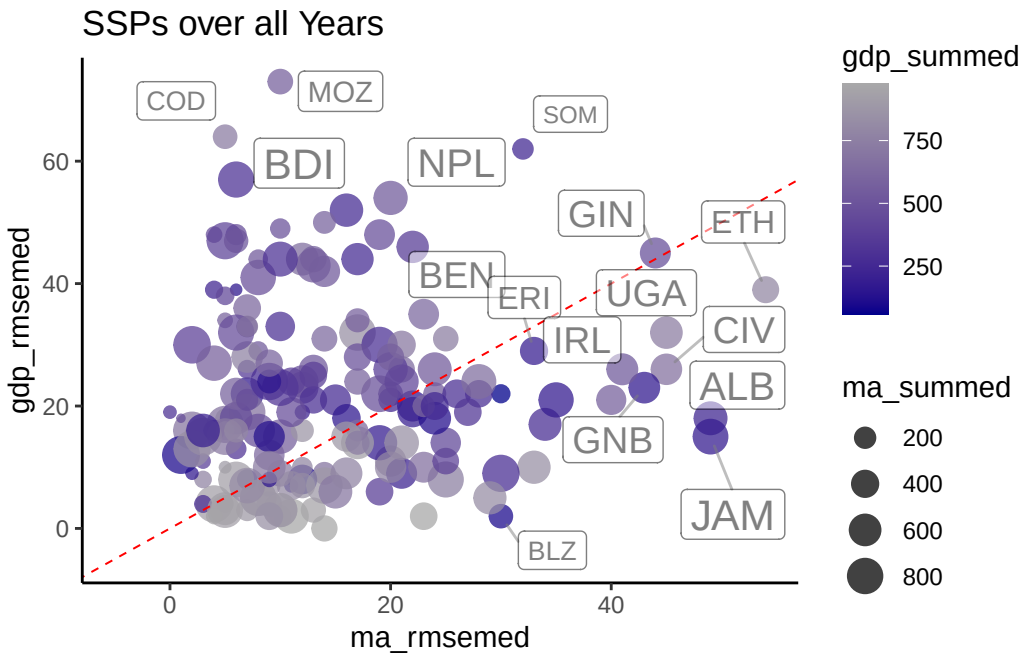



```
out_2100 %>%
  drop_na() %>%
  ggplot(aes(x=health_rng, y=gdp_rng, size=health_med, colour=gdp_med)) +
  p2100
```

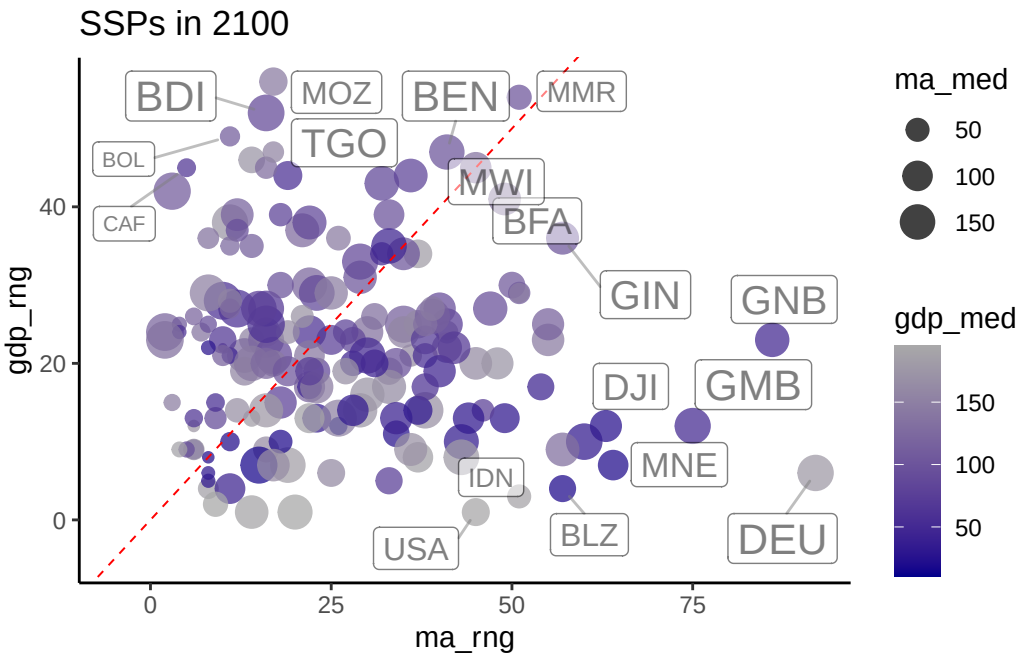


Market Access

```
out_all %>%
  drop_na() %>%
  ggplot(aes(x=ma_rmsemed, y=gdp_rmsemed, size=ma_summed, colour=gdp_summed)) +
  pall
```

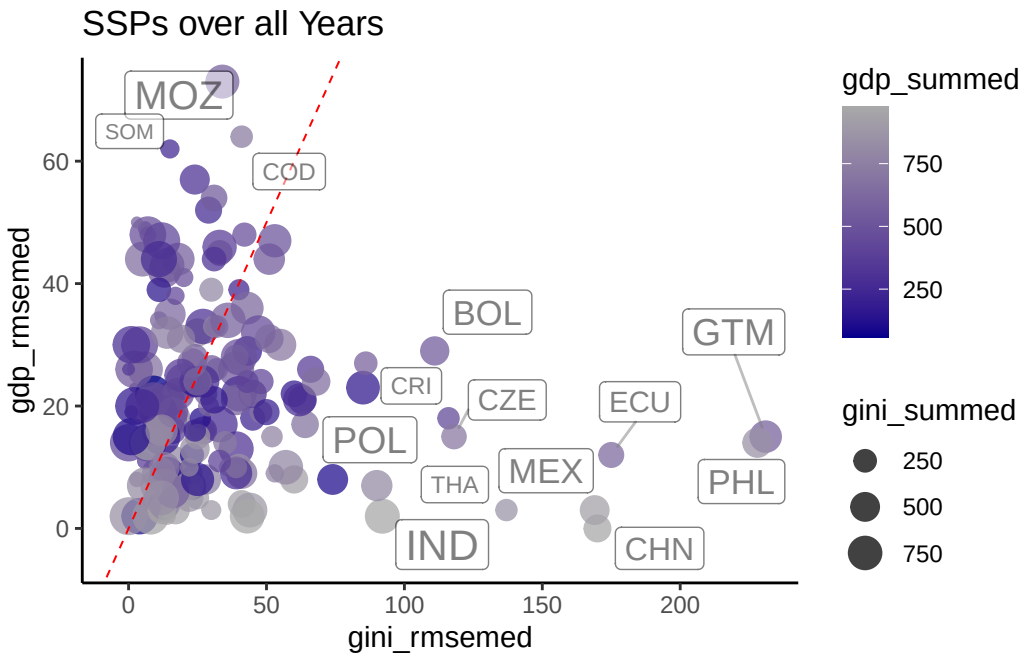



```
out_2100 %>%
  drop_na() %>%
  ggplot(aes(x=ma_rng, y=gdp_rng, size=ma_med, colour=gdp_med)) +
  p2100
```

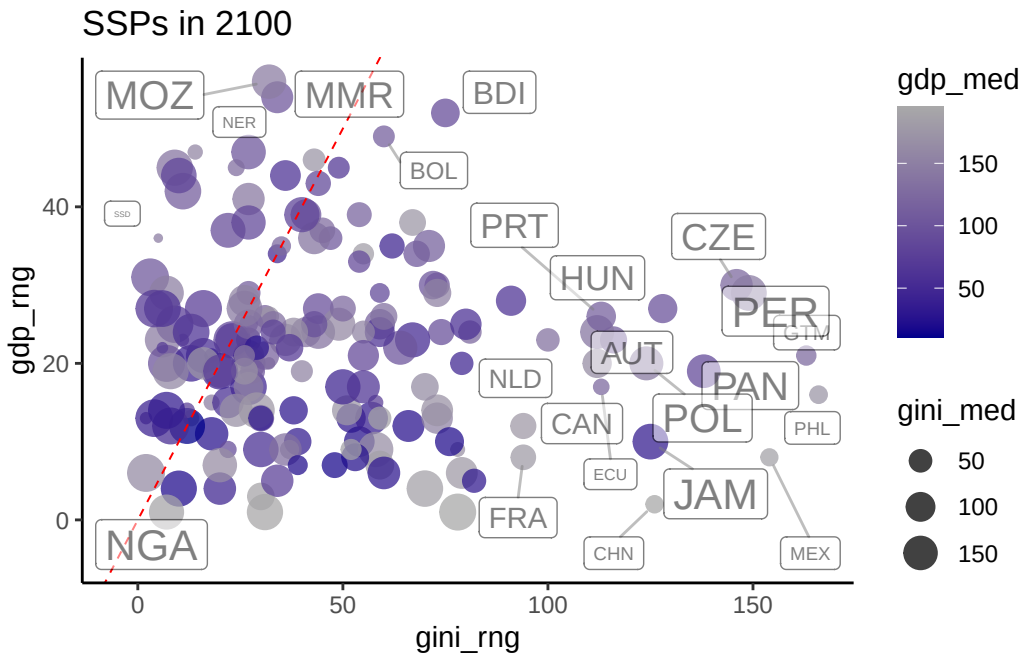


Gini

```
out_all %>%
  drop_na() %>%
  ggplot(aes(x=gini_rmsemed, y=gdp_rmsemed, size=gini_summed, colour=gdp_summed)) +
  pall
```

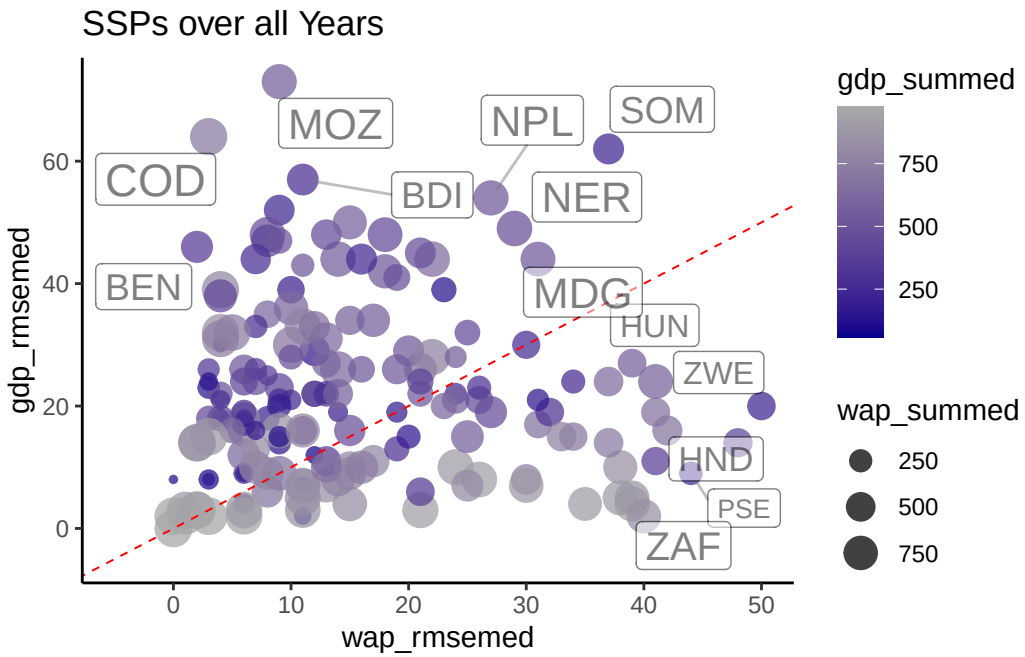



```
out_2100 %>%
  drop_na() %>%
  ggplot(aes(x=gini_rng, y=gdp_rng, size=gini_med, colour=gdp_med)) +
  p2100
```

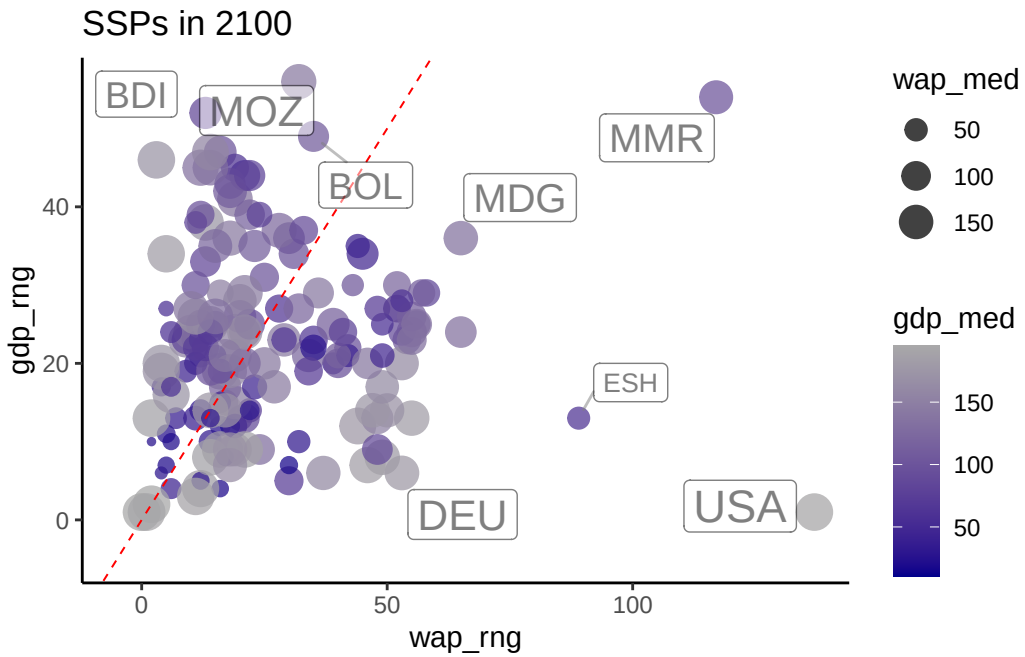


Working Age Population

```
out_all %>%
  drop_na() %>%
  ggplot(aes(x=wap_rmsemed, y=gdp_rmsemed, size=wap_summed, colour=gdp_summed)) +
  pall
```

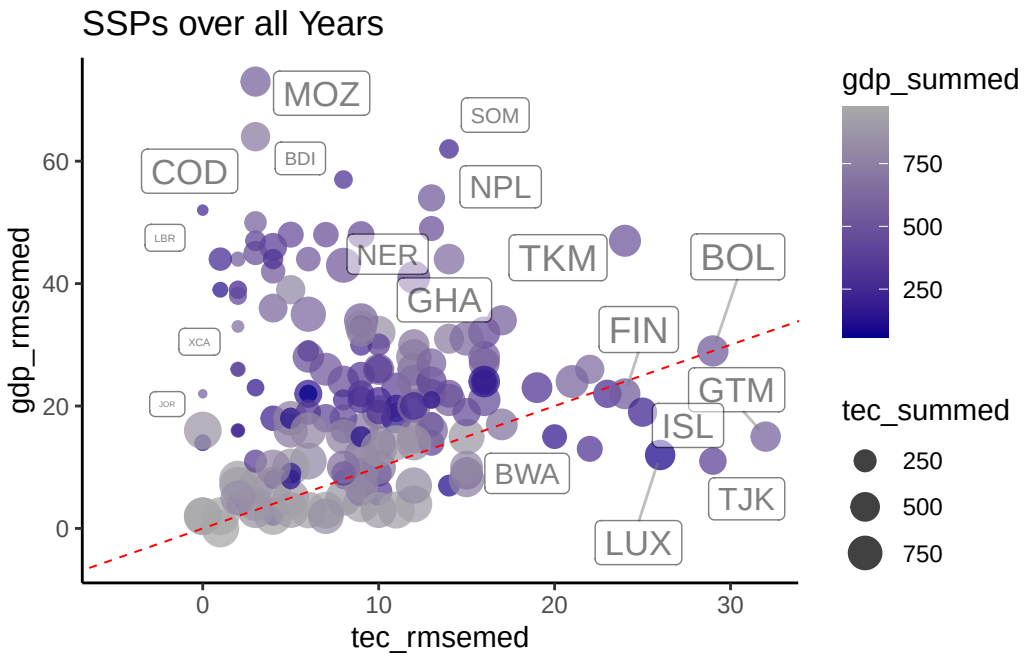



```
out_2100 %>%
  drop_na() %>%
  ggplot(aes(x=wap_rng, y=gdp_rng, size=wap_med, colour=gdp_med)) +
  p2100
```

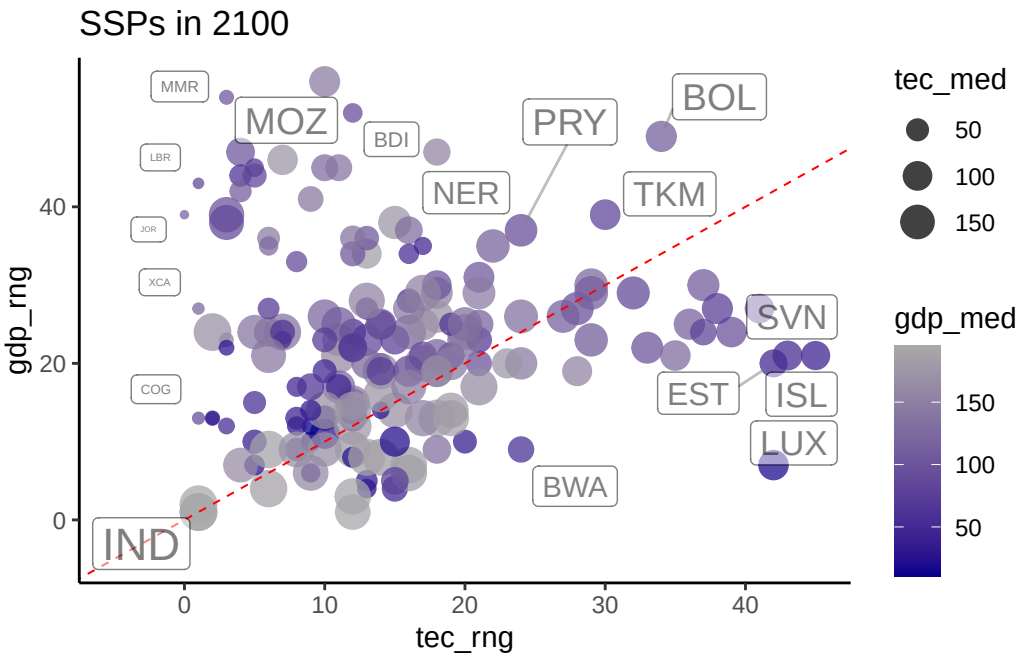


Energy

```
out_all %>%
  drop_na() %>%
  ggplot(aes(x=tec_rmsemed, y=gdp_rmsemed, size=tec_summed, colour=gdp_summed)) +
  pall
```




```
out_2100 %>%
  drop_na() %>%
  ggplot(aes(x=tec_rng, y=gdp_rng, size=tec_med, colour=gdp_med)) +
  p2100
```



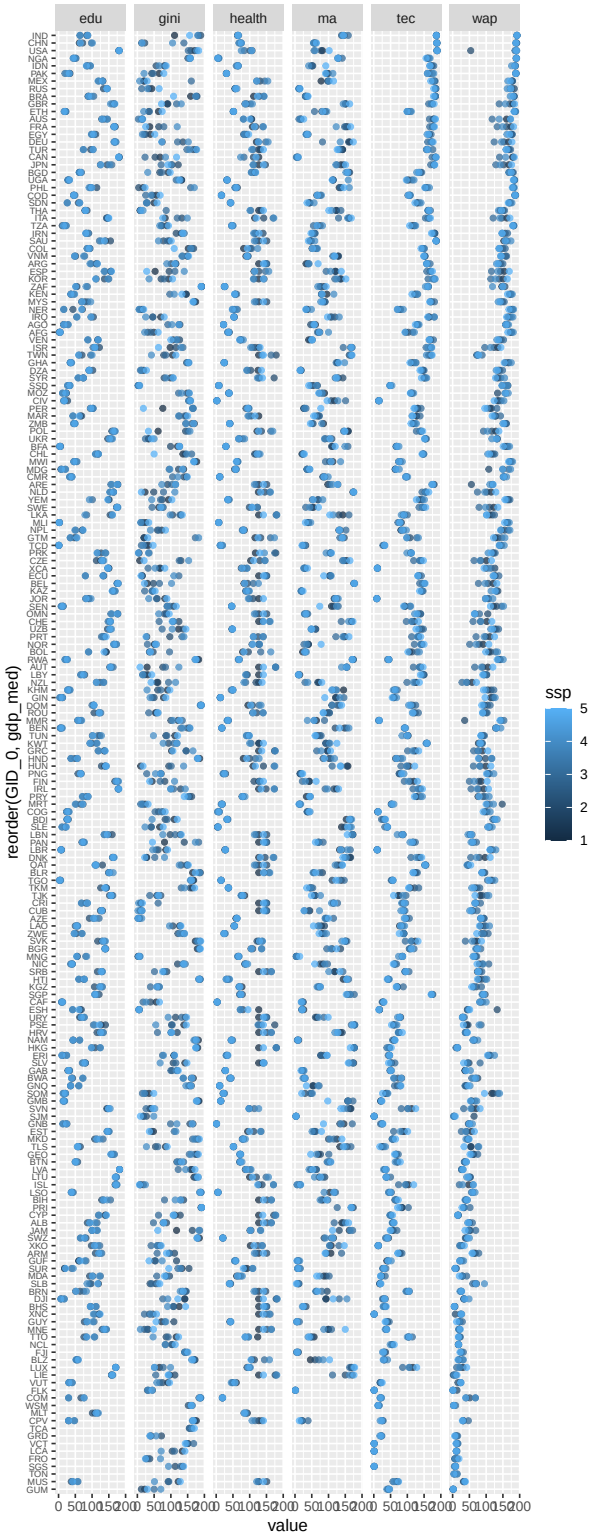
Experimental Another way to plot these (if GDP is always the comparator) might be to show countries ranked on y-axis by gdp (med rank) then points on xaxis for the other variable value (for each ssp), then facet on x for variables (but this then does not show countries with variable vs non-variable gdp)

However, the below shows that with 200+ countries it's hard to get this looking good. It does show a string relationship between GDP and WAP (and to some degree energy).

```
#create data for 2100 to trial gdp ranked plot

ranks_2100 = edu
counter = 1
for(nm in list(edu,ma, health, gini, wap, tec)){

  if(counter == 1){
    ranks_2100 <- filter(nm, unit=='rank', year==2100)
  } else {
    ranks_2100 <- bind_rows(ranks_2100, filter(nm, unit=='rank', year==2100))
  }
  counter = counter + 1
}
```

```r
#create gdp rank data
gdp_2100 <-
  out_2100 %>%
  select(OID, GID_0, gdp_med)

#join ranks for our metrics to gdp
gdp_ranks_2100 <-
  left_join(gdp_2100, ranks_2100, by='GID_0',suffix=c("",".y")) %>%
  select(-ends_with(".y"))
```

```r
#plot for EDU
gdp_ranks_2100 %>%
  drop_na() %>%
  arrange(gdp_med) %>%  #order countries by median SSP GDP
  #filter(variable=='edu') %>%
  ggplot(aes(x=value, y=reorder(GID_0,gdp_med), colour=ssp)) +
  geom_point(alpha=0.75) +
  theme(axis.text.y = element_text(size=rel(0.65))) +
  facet_grid(.~variable)
```


edu
gini
health
ma
tec
wap
reorder(GID_0, gdp_med)
value
ssp
5
4
3
2
1