

Manipulação e Apresentação de Dados

Guilherme Bovi Ambrosano

Exemplos

Conjuntos de dados do governo



<https://dados.gov.br/dados/conjuntos-dados>

Plantas alternativas a exóticas invasoras



Fonte: <https://dados.gov.br/dados/conjuntos-dados/plantas-alternativas-as-plantas-exoticas-invasoras-rj>

```
library(readxl)
read_xlsx(
  paste0("../dados/listagem-de-plantas-alternativas-as-",
          "plantas-exoticas-invasoras-listadas-para-o-",
          "estado-do-rio-de.xlsx"),
  range="A7:L268")

## New names:
## * `Nome científico` -> `Nome científico...5`
## * `Nativa (N) ou Exótica (E)` -> `Nativa (N) ou Exótica (E)..`
## * `Nome científico` -> `Nome científico...7`
## * `Nativa (N) ou Exótica (E)` -> `Nativa (N) ou Exótica (E)..`
## * `Nome científico` -> `Nome científico...9`
## * `Nativa (N) ou Exótica (E)` -> `Nativa (N) ou Exótica (E)..`

## # A tibble: 261 x 12
##   `Código de uso` `Família EEI` `Nome científico EEI`
##   <chr>          <chr>          <chr>
```



Plantas alternativas a exóticas invasoras

Fonte: <https://dados.gov.br/dados/conjuntos-dados/plantas-alternativas-as-plantas-exoticas-invasoras-rj>

```
library(readxl)
invasoras <- read_xlsx(
  paste0("../dados/listagem-de-plantas-alternativas-as-",
    "plantas-exoticas-invasoras-listadas-para-o-",
    "estado-do-rio-de.xlsx"),
  col_names=c("Codigo", "Familia_EEI", "NomeCient_EEI",
    "NomePop_EEI",
    "NomeCient_Ombro", "NouE_Ombro",
    "NomeCient_Semidec", "NouE_Semidec",
    "NomeCient_InfMar", "NouE_InfMar",
    "Ecossist_EEI", "Categoria"),
  range="A8:L268")
```

Plantas alternativas a exóticas invasoras



```
invasoras %>%  
  pivot_longer(c(ends_with("Ombro"),  
                 ends_with("Semidec"),  
                 ends_with("InfMar"))) %>%  
  separate(name, c("Var", "Ecosystem"))
```

```
## # A tibble: 1,566 x 9
```

```
##   Codigo      Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem  
##   <chr>      <chr>      <chr>      <chr>      <chr>  
## 1 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 2 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 3 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 4 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 5 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 6 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 7 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 8 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 9 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 10 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações
```

```
## #> # A tibble: 1,566 x 9
```

Plantas alternativas a exóticas invasoras



```
invasoras %>%
  pivot_longer(c(ends_with("Ombro"),
                 ends_with("Semidec"),
                 ends_with("InfMar")) %>%
  separate(name, c("Var", "Ecosystem")) %>%
  pivot_wider(id_cols=c(Codigo,ends_with("EEI"),Categoria,Ecosystem),
              names_from=Var)
```

```
## # A tibble: 354 x 9
```

```
##   Codigo  Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem_EEI Categoria
##   <chr>   <chr>         <chr>         <chr>         <chr>         <chr>
## 1 orname~ Acanthaceae Asystasia ga~ assistásia,~ Formações P~
## 2 orname~ Acanthaceae Asystasia ga~ assistásia,~ Formações P~
## 3 orname~ Acanthaceae Asystasia ga~ assistásia,~ Formações P~
## 4 orname~ Acanthaceae Brillantaisi~ erva-de-bi~ Floresta Om~
## 5 orname~ Acanthaceae Brillantaisi~ erva-de-bi~ Floresta Om~
## 6 orname~ Acanthaceae Brillantaisi~ erva-de-bi~ Floresta Om~
## 7 orname~ Acanthaceae Thunbergia a~ amarelinha~ Floresta Om~
## 8 orname~ Acanthaceae Thunbergia a~ amarelinha~ Floresta Om~
## 9 orname~ Acanthaceae Thunbergia a~ amarelinha~ Floresta Om~
## 10 orname~ Acanthaceae Thunbergia g~ tumbérgia-- Floresta Om~
```

Plantas alternativas a exóticas invasoras



```
invasoras %>%
  pivot_longer(c(ends_with("Ombro"),
                 ends_with("Semidec"),
                 ends_with("InfMar"))) %>%
  separate(name, c("Var", "Ecosystem")) %>%
  pivot_wider(id_cols=c(Codigo,ends_with("EEI"),Categoria,Ecosystem),
              names_from=Var) %>%
  unnest_longer(c(NomeCient, NouE))
```

```
## # A tibble: 783 x 9
```

```
##   Codigo  Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem_EEI Categoria
##   <chr>   <chr>         <chr>         <chr>         <chr>         <chr>
## 1 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 2 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 3 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 4 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 5 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 6 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 7 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 8 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 9 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
```


Plantas alternativas a exóticas invasoras



```
invasoras %>%
  pivot_longer(c(ends_with("Ombro"),
                 ends_with("Semidec"),
                 ends_with("InfMar"))) %>%
  separate(name, c("Var", "Ecosystem")) %>%
  pivot_wider(id_cols=c(Codigo,ends_with("EEI"),Categoria,Ecosystem),
              names_from=Var) %>%
  unnest_longer(c(NomeCient, NouE)) %>%
  filter(!is.na(NomeCient), !is.na(NouE))
```

```
## # A tibble: 504 x 9
```

```
##   Codigo  Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem_EEI Categoria
##   <chr>   <chr>         <chr>         <chr>         <chr>         <chr>
## 1 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 2 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 3 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 4 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 5 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 6 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 7 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 8 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
```



```
group_by(NomeCient_EEI) %>%
```

count() %>%

```
ungroup() %>%
```

```
ggplot(aes(x=NomeCient_EEI, y=n)) +
```

geom_col() +

```
theme_grey(base_size = 8) +
```

```
theme(axis.text.x = element_text(angle=90, hjust=1, vjust=.5))
```





```
theme(axis.text.x = element_text(angle=90, hjust=1, vjust=.5))
```



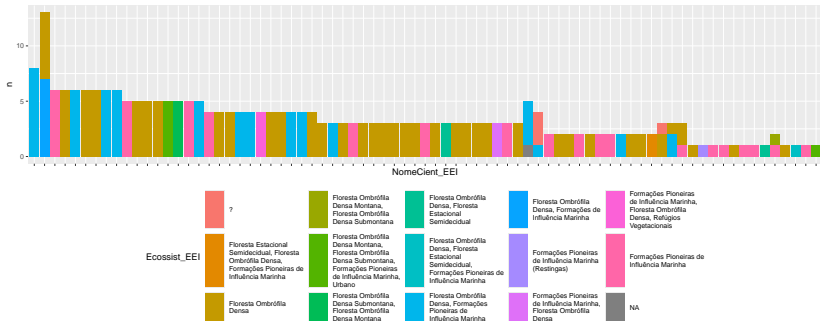


Plantas alternativas a exóticas invasoras

```

invasoras %>%
  group_by(NomeCient_EEI, Ecosystem_EEI) %>%
  count() %>%
  ungroup() %>%
  mutate(NomeCient_EEI = fct_reorder(NomeCient_EEI, desc(n))) %>%
  mutate_at(vars(Ecosystem_EEI), str_wrap, width=22) %>%
  ggplot(aes(x=NomeCient_EEI, y=n, fill=Ecosystem_EEI)) +
  guides(fill=guide_legend(nrow=3)) +
  geom_col() +
  theme_grey(base_size = 8) +
  theme(axis.text.x = element_blank(),
        legend.position = "bottom")

```



Plantas alternativas a exóticas invasoras



```
invasoras %>%  
  group_by(NomeCient_EEI, Ecossist_EEI) %>%  
  count() %>%  
  group_by(NomeCient_EEI) %>%  
  count() %>%  
  arrange(desc(n))
```

```
## # A tibble: 77 x 2  
## # Groups:   NomeCient_EEI [77]  
##   NomeCient_EEI      n  
##   <chr>          <int>  
## 1 Acacia angustissima (Mill.) Kuntze      2  
## 2 Acacia mangium Wild                    2  
## 3 Dieffenbachia amoena hort. ex Gentil    2  
## 4 Elaeis guineensis Jacq.                 2  
## 5 Melinis minutiflora P. Beauv.           2  
## 6 Mimusops commersonii (G. Don) Engl.     2  
## 7 Thunbergia grandiflora Roxb.            2  
## 8 Acacia auriculiformis A.Cunn. ex Benth.  1  
## 9 Aroca americana L.                     1
```



Plantas alternativas a exóticas invasoras

```
ecossist <- invasoras %>%  
  mutate_at(vars(Ecossist_EEI), replace_na, "?") %>%  
  group_by(NomeCient_EEI, Ecossist_EEI) %>%  
  count() %>%  
  group_by(NomeCient_EEI) %>%  
  arrange(Ecossist_EEI) %>%  
  summarise(Ecossist_EEI = (function(x) {  
    if(length(x)==1) return(x)  
    else {  
      if(is.na(x[1])) return(x[2])  
      if(is.na(x[2])) return(x[1])  
      if(grepl(x[1], x[2])) return(x[2])  
      if(grepl(x[2], x[1])) return(x[1])  
      return(paste(x, collapse=", "))  
    }  
  }  
) (Ecossist_EEI),  
  n = sum(n)) %>%  
  mutate_at(vars(Ecossist_EEI), ~ifelse(  
    .=="Formações Pioneiras de Influência Marinha, Floresta Ombrófila D  
    "Floresta Ombrófila Densa, Formações Pioneiras de Influência Marinh
```

Plantas alternativas a exóticas invasoras



```
ecossist %>%
  mutate(NomeCient_EEI = fct_reorder(NomeCient_EEI, desc(n))) %>%
  mutate_at(vars(Ecossist_EEI), str_wrap, width=22) %>%
  ggplot(aes(x=NomeCient_EEI, y=n, fill=Ecossist_EEI)) +
  guides(fill=guide_legend(nrow=3)) +
  geom_col() +
  theme_grey(base_size = 8) +
  theme(axis.text.x = element_blank(),
        legend.position = "bottom")
```





Fonte: <https://dados.gov.br/dados/conjuntos-dados/programa-fomento-rural>

```
library(readr)
```

```
fomentoRural <- read_csv(  
  paste0("https://aplicacoes.mds.gov.br/sagi/",  
        "servicos/misocial/?",  
        "fq=anomes_s:2023*&",  
        "fq=formento_qtd_total_familias_benef_i:*&",  
        "q=:*&",  
        "rows=1000000&",  
        "wt=csv"))
```

```
head(fomentoRural)
```

```
## # A tibble: 6 x 1,077
```

```
##   paa_qtd_agricul_familiar_modal~1 qtd_pes_pbf_idade_7_~2 bolsa_atle
```

```
##   <lgl>                                     <dbl> <lgl>
```

```
## 1 NA                                     148 NA
```

```
## 2 NA                                     82 NA
```

```
## 3 NA                                     118 NA
```

```
## 4 NA                                     128 NA
```




```
fomentoRural %>%  
  select_if(~!all(is.na(.))) %>%  
  select_if(~!all(==0))
```

```
## # A tibble: 662 x 367
```

```
##   qtd_pes_pbf_idade_7_a_15_sexo~1 cadunico_tot_fam_ext~2 igd_pbf_qt
```

```
##                                     <dbl>                 <dbl>
```

```
## 1                                148                     NA
```

```
## 2                                82                     4030
```

```
## 3                                118                    7707
```

```
## 4                                128                     NA
```

```
## 5                                 79                     NA
```

```
## 6                                 85                    5489
```

```
## 7                                 83                    5522
```

```
## 8                                 24                     756
```

```
## 9                                 81                    3814
```

```
## 10                               189                    2728
```

```
## # i 652 more rows
```

```
## # i abbreviated names: 1: qtd_pes_pbf_idade_7_a_15_sexo_feminino_i,
```

```
## #   2: cadunico_tot_fam_ext_pob_i, 3: igd_pbf_qtd_total_publico_saud
```

```
## # i 364 more variables:
```



```
fomentoRural %>%  
  select(anomes_s, matches( paste0(  
    "formento\\_qtd\\_fam\\_benef\\_.*\\_i")))
```

```
## # A tibble: 662 x 15
```

```
##   anomes_s formento_qtd_fam_ben~1 formento_qtd_fam_ben~2 for
```

```
##           <dbl>                  <dbl>                  <dbl>
```

```
## 1    202312                      0                      0
```

```
## 2    202311                      0                      0
```

```
## 3    202310                      0                      0
```

```
## 4    202312                      1                      0
```

```
## 5    202312                      0                      0
```

```
## 6    202310                      0                      0
```

```
## 7    202311                      0                      0
```

```
## 8    202306                      0                      0
```

```
## 9    202301                      0                      0
```

```
## 10   202301                      0                      0
```

```
## # i 652 more rows
```

```
## # i abbreviated names:
```

```
## # 1: formento_qtd_fam_benef_catadones_material_especializado_i
```



```
fomentoRural %>%  
  select(anomes_s, matches(  
    "formento\\_qtd\\_fam\\_benef\\_.*\\_i")) %>%  
  group_by(anomes_s) %>%  
  summarise_all(sum, na.rm=T)
```

```
## # A tibble: 10 x 15  
##   anomes_s formento_qtd_fam_ben~1 formento_qtd_fam_ben~2 formento_q  
##   <dbl>          <dbl>          <dbl>  
## 1  202301            0            0  
## 2  202303            0            0  
## 3  202304            0            0  
## 4  202306            1            0  
## 5  202307            0            0  
## 6  202308            0            0  
## 7  202309            0            0  
## 8  202310            0            0  
## 9  202311            5            0  
## 10 202312            6            1  
## # i abbreviated names:  
## # 1: formento_qtd_fam_benef_catadores_material_reciclado_i,
```



```
fr <- fomentoRural %>%
  select(anomes_s, matches(
    "formento\\_qtd\\_fam\\_benef\\_.*\\_i")) %>%
  group_by(anomes_s) %>%
  summarise_all(sum, na.rm=T) %>%
  pivot_longer(-anomes_s)

head(fr)
```

```
## # A tibble: 6 x 3
##   anomes_s name                                     val
##   <dbl> <chr>                                     <dbl>
## 1  202301 formento_qtd_fam_benef_catadores_material_reciclado_i
## 2  202301 formento_qtd_fam_benef_comun_terreiro_i
## 3  202301 formento_qtd_fam_benef_ciganas_i
## 4  202301 formento_qtd_fam_benef_quilombolas_i
## 5  202301 formento_qtd_fam_benef_assentados_ref_agraria_i
## 6  202301 formento_qtd_fam_benef_agrilcultores_familiares_i      2
```



```
fr <- fr %>%  
  mutate_at(vars(anomes_s), as.character) %>%  
  mutate_at(vars(anomes_s), fast_strptime, format="%Y%m", lt=F)  
  
head(fr)
```

```
## # A tibble: 6 x 3  
##   anomes_s          name  
##   <dtm>           <chr>  
## 1 2023-01-01 00:00:00 formento_qtd_fam_benef_catadores_material_recib  
## 2 2023-01-01 00:00:00 formento_qtd_fam_benef_comun_terreiro_i  
## 3 2023-01-01 00:00:00 formento_qtd_fam_benef_ciganas_i  
## 4 2023-01-01 00:00:00 formento_qtd_fam_benef_quilombolas_i  
## 5 2023-01-01 00:00:00 formento_qtd_fam_benef_assentados_ref_agraria_  
## 6 2023-01-01 00:00:00 formento_qtd_fam_benef_agricultores_familiare
```



```
fr <- fr %>%  
  mutate(GPTE = str_remove(name, "formento\\_qtd\\_fam\\_benef\\_"),  
         GPTE = str_remove(GPTE, "\\_i$")) %>%  
  mutate_at(vars(GPTE), factor)
```

```
head(fr)
```

```
## # A tibble: 6 x 4
```

```
##   anomes_s      name
```

```
##   <dtm>      <chr>
```

```
## 1 2023-01-01 00:00:00 formento_qtd_fam_benef_catadores_material_rec~
```

```
## 2 2023-01-01 00:00:00 formento_qtd_fam_benef_comun_terreiro_i
```

```
## 3 2023-01-01 00:00:00 formento_qtd_fam_benef_ciganas_i
```

```
## 4 2023-01-01 00:00:00 formento_qtd_fam_benef_quilombolas_i
```

```
## 5 2023-01-01 00:00:00 formento_qtd_fam_benef_assentados_ref_agraria~
```

```
## 6 2023-01-01 00:00:00 formento_qtd_fam_benef_agrilcultores_familiar~
```



```
fr <- fr %>%  
  mutate_at(vars(GPTE), fct_relabel,  
             str_replace_all, "\\_", " ") %>%  
  mutate_at(vars(GPTE), fct_relabel,  
             str_to_title)
```

```
levels(fr$GPTE)
```

```
## [1] "Acampados Rurais"           "Agricultores Familiares"  
## [3] "Assentados Ref Agraria"     "Atingidas Empr Infra"  
## [5] "Catadores Material Reciclado" "Ciganas"  
## [7] "Comun Terreiro"            "Credito Fundiario"  
## [9] "Extrativistas"             "Indinegas"  
## [11] "Pescadores Artesanais"     "Presos Sist Carcerario"  
## [13] "Quilombolas"               "Ribeirinhas"
```

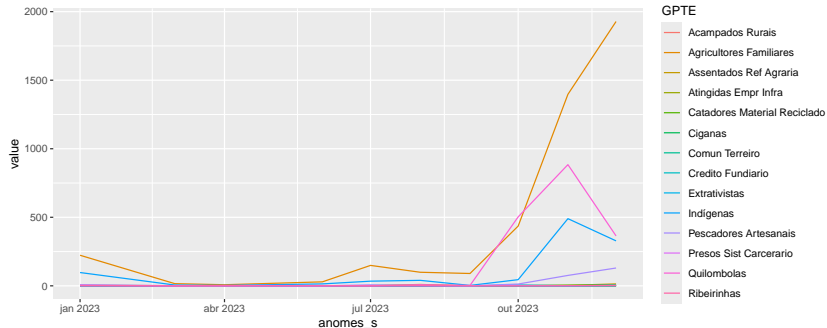


```
fr <- fr %>%  
  mutate_at(vars(GPTE), fct_relabel,  
             str_replace,  
             "Agrilcultores", "Agricultores") %>%  
  mutate_at(vars(GPTE), fct_relabel,  
             str_replace,  
             "Indinegas", "Indígenas")  
  
levels(fr$GPTE)
```

```
## [1] "Acampados Rurais"           "Agricultores Familiares"  
## [3] "Assentados Ref Agraria"     "Atingidas Empr Infra"  
## [5] "Catadores Material Reciclado" "Ciganas"  
## [7] "Comun Terreiro"            "Credito Fundiario"  
## [9] "Extrativistas"             "Indígenas"  
## [11] "Pescadores Artesanais"     "Presos Sist Carcerario"  
## [13] "Quilombolas"               "Ribeirinhas"
```




```
fr %>%  
  ggplot(aes(x=anomes_s, col=GPTE, y=value)) +  
  geom_line()
```



Ranking de competitividade por estados



<https://tinyurl.com/mry3wrp7>

```
indicadores <- read_xlsx("../dados/ranking_2015_2020.xlsx")  
head(indicadores)
```

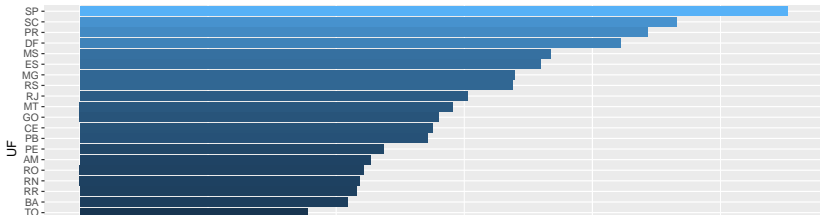
```
## # A tibble: 6 x 15
```

```
##   Pilar                               Indicador  Ordem UF      `Ano Atualizaçã  
##   <chr>                               <chr>      <dbl> <chr>      <db  
## 1 Potencial de Mercado Tamanho d~      1 AC      20  
## 2 Potencial de Mercado Taxa de c~      1 AC      20  
## 3 Potencial de Mercado Total          1 AC      20  
## 4 Infraestrutura      Acessibil~      1 AC      20  
## 5 Infraestrutura      Custo de ~     -1 AC      20  
## 6 Infraestrutura      Custo de ~     -1 AC      20  
## # i 9 more variables: `Nota Bruta Indicador` <chr>, `Peso Ind  
## #   `Peso Pilares` <lgl>, `Peso Real` <lgl>, `Descrição Sintá  
## #   `Fonte Oficial` <chr>, `Unidade de Medida` <chr>, Detalhe  
## #   `Nota Normalizada Indicador` <dbl>
```

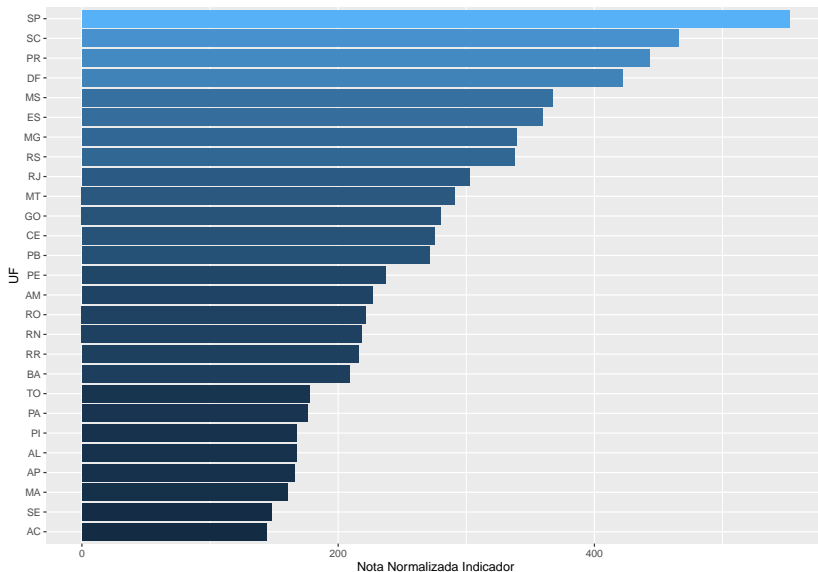


Ranking de competitividade por estados

```
indicadores %>%  
  filter(Pilar=="Ranking Geral",  
         Indicador=="Total",  
         str_ends(UF, "[^*]")) %>%  
  mutate(UF=fct_reorder(UF, `Nota Normalizada Indicador`,  
                        sum, .na_rm=T)) %>%  
  group_by(UF) %>%  
  summarise_at(vars(`Nota Normalizada Indicador`,  
                   sum, na.rm=T) %>%  
ggplot(aes(x=`Nota Normalizada Indicador`, y=UF)) +  
geom_col(aes(fill=`Nota Normalizada Indicador`,  
              show.legend=F)
```



Ranking de competitividade por estados



Ranking de competitividade por estados



```
library(geobr)
```

```
estados <- read_state(  
  year = 2020,  
  showProgress = FALSE)
```

```
estados
```

```
## Simple feature collection with 27 features and 5 fields
```

```
## Geometry type: MULTIPOLYGON
```

```
## Dimension:      XY
```

```
## Bounding box:  xmin: -73.99045 ymin: -33.75118 xmax: -28.8476
```

```
## Geodetic CRS:  SIRGAS 2000
```

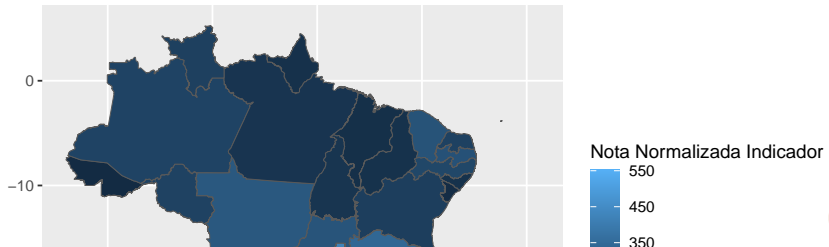
```
## First 10 features:
```

##	code_state	abbrev_state	name_state	code_region	name_region
## 1	11	RO	Rondônia	1	Norte
## 2	12	AC	Acre	1	Norte
## 3	13	AM	Amazônas	1	Norte
## 4	14	RR	Roraima	1	Norte
## 5	15	PA	Pará	1	Norte

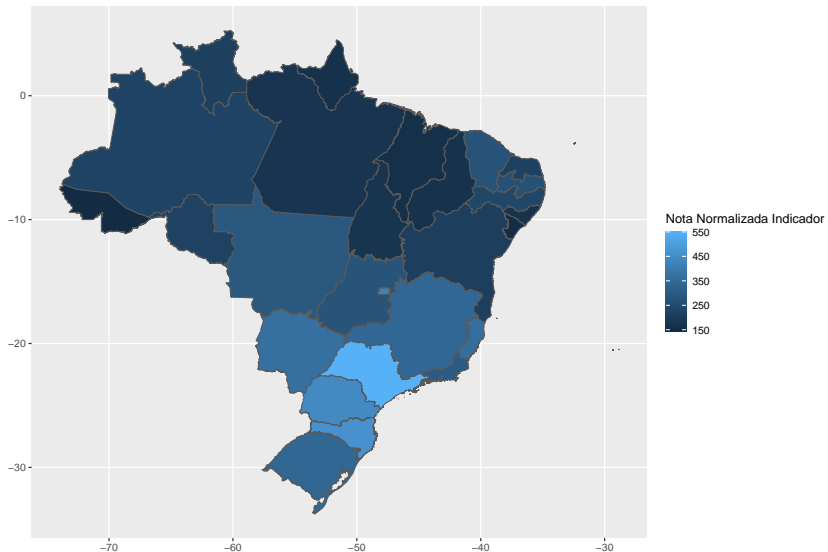


Ranking de competitividade por estados

```
indicadores %>%  
  filter(Pilar=="Ranking Geral",  
         Indicador=="Total",  
         str_ends(UF, "[^*]")) %>%  
  group_by(UF) %>%  
  summarise_at(vars(`Nota Normalizada Indicador`),  
               sum, na.rm=T) %>%  
  rename(abbrev_state=UF) %>%  
  left_join(estados, by="abbrev_state") %>%  
  ggplot(aes(fill=`Nota Normalizada Indicador`, geometry=geom))  
  geom_sf()
```



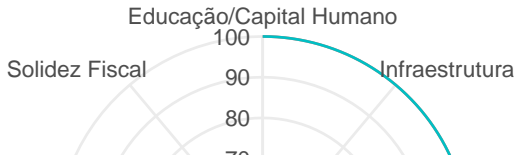
Ranking de competitividade por estados





Ranking de competitividade por estados

```
indicadores %>%  
  filter(Pilar!="Ranking Geral",  
         Indicador=="Total", UF=="SP",  
         `Ano Publicação` %in% c(2019, 2020)) %>%  
  group_by(`Ano Publicação`, Pilar) %>%  
  summarise(soma = sum(`Nota Normalizada Indicador`)) %>%  
  mutate(`Ano Publicação` = factor(`Ano Publicação`)) %>%  
  mutate(Pilar = fct_reorder(Pilar, -soma)) %>%  
  mutate(Pilar = fct_relabel(Pilar, str_wrap, width=15)) %>%  
  ggplot(aes(x=Pilar, y=soma, color=`Ano Publicação`)) +  
  geom_line(aes(group=`Ano Publicação`)) +  
  coord_radial(expand=F, r_axis_inside = T) +  
  theme_minimal() +  
  theme(axis.title = element_blank())
```



Ranking de competitividade por estados

