

# Cadernos de Scripts - Módulo 2

Daniel Pagotto e Daiane Martins (LAPEI-UFG)

Esse é o caderno de Scripts do módulo 2, no qual aplicaremos alguns conceitos de manipulação e visualização de dados do módulo 1 na base do Global Entrepreneurship Monitor (GEM). Antes de iniciar é importante assegurar que os seguintes pacotes estejam instalados: `readr`, `dplyr`, `skimr`, `ggplot2`, `ggrepel`, `directlabels`, `GGally`. Caso não tenha ainda, use `install.packages()` para baixá-los.

Vamos carregar os pacotes que vamos usar ao longo da aula!

## Sobre o GEM

Desde 1999, o GEM realiza levantamentos para monitorar o empreendedorismo em diferentes países do mundo. Para isso, pesquisas com empreendedores e especialistas são anualmente realizadas para mapear elementos como o contexto empreendedor, comportamentos e atitudes empreendedoras.

O GEM possui duas bases essencialmente: 1) National Expert Survey (especialistas); 2) Adult Population Survey. Os microdados são acessíveis após três anos da sua aplicação. Porém, conseguimos acessar dados consolidados a partir do site do GEM. Para mais informações sobre os levantamentos, acesso aos dados e a relatórios de cada país, acesse a página do GEM.

## National Expert Survey (NES)

A base NES do GEM contempla diferentes dimensões:

- **Finanças empreendedoras:** disponibilidade de diferentes fontes de finanças para pequenas e médias empresas (PMEs).
- **Políticas governamentais:** o quanto as políticas públicas contribuem para o empreendedorismo:
  - O quanto empreendedorismo é importante como uma política de governo
  - O quanto impostos e burocracia afetam empreendedorismo
- **Programas governamentais para empreendedorismo:** presença de programas governamentais para empreendedorismo.
- **Educação empreendedora:** presença de treinamentos sobre criação e gerenciamento de PMEs em diferentes graus de ensino:
  - Nível básico
  - Nível Superior
- **Transferência de Pesquisa & Desenvolvimento (P&D):** o quanto que pesquisas nacionais conduzem a criação de oportunidades comerciais e estão disponíveis para PMEs.
- **Infraestrutura legal e comercial:** presença de serviços e instituições que fornecem suporte a PMEs.

- **Regulação de entrada:** que é dividido em a) dinamismo de mercado - o quanto que o mercado muda de ano a ano; b) abertura de mercado - o quanto novas empresas são livres para entrar em mercados existentes.
- **Infraestrutura física:** facilidade de acessar recursos físicos (ex.: comunicação, logística, espaço físico).
- **Cultura e normas sociais:** o quanto normas sociais e culturais encorajam novos métodos de negócios que podem gerar riqueza e renda de uma pessoa.

Vamos fazer alguns estudos

## Lendo a base a partir do meu repositório do github

```
gem_especialistas <- read_delim("https://raw.githubusercontent.com/danielppagotto/R_empreendedorismo2/main/gem_especialistas.csv")
```

```
##
## -- Column specification -----
## cols(
##   code = col_double(),
##   economy = col_character(),
##   year = col_double(),
##   financing = col_double(),
##   gov_support = col_double(),
##   taxes_bureaucracy = col_double(),
##   gov_programs = col_double(),
##   basic_school_entrepreneurship = col_double(),
##   post_school_entrepreneurship = col_double(),
##   rd_transfer = col_double(),
##   commercial_professional_infraestructure = col_double(),
##   internal_market_dynamics = col_double(),
##   internal_market_openness = col_double(),
##   physical_infrastructure = col_double(),
##   cultural_social_norms = col_double()
## )
```

## Compreendendo as variáveis da base

Vamos usar o `glimpse` para inspecionar os dados na base.

```
glimpse(gem_especialistas)
```

```
## Rows: 649
## Columns: 15
## $ code                <dbl> 244, 43, 55, 226, 56, 57, 385, ~
## $ economy             <chr> "Angola", "Austria", "Brazil", ~
## $ year                <dbl> 2020, 2020, 2020, 2020, 2020, 2~
## $ financing           <dbl> 2.18, 2.88, 2.73, 1.95, 2.17, 2~
## $ gov_support         <dbl> 2.50, 2.79, 2.36, 2.16, 2.54, 2~
## $ taxes_bureaucracy   <dbl> 1.98, 2.53, 1.85, 2.34, 2.78, 2~
## $ gov_programs        <dbl> 2.00, 3.60, 2.56, 2.35, 3.02, 2~
```

```
## $ basic_school_entrepreneurship      <dbl> 1.93, 1.74, 1.94, 1.55, 1.82, 2~
## $ post_school_entrepreneurship      <dbl> 2.20, 2.67, 2.73, 2.25, 2.81, 3~
## $ rd_transfer                       <dbl> 1.74, 2.77, 2.16, 2.04, 2.40, 2~
## $ commercial_professional_infraestructure <dbl> 2.14, 3.27, 2.79, 2.89, 2.82, 2~
## $ internal_market_dynamics          <dbl> 3.07, 2.70, 3.73, 2.47, 2.66, 2~
## $ internal_market_openess          <dbl> 2.09, 3.34, 2.55, 2.25, 2.23, 2~
## $ physical_infrastructure           <dbl> 2.23, 4.20, 3.40, 2.76, 3.99, 3~
## $ cultural_social_norms             <dbl> 2.85, 2.63, 2.89, 2.60, 3.05, 3~
```

## Explorando base de especialistas

Vamos baixar a base e comparar alguns países: **Brasil, Chile, Colômbia e México.**

```
países <- c("Brazil", "Chile", "Colombia", "Mexico")
```

Além disso, vamos separar algumas variáveis de interesse: `economy`, `year`, `gov_support`, `taxes_bureaucracy`, `gov_programs`, `internal_market_dynamics`.

```
gem_países_selecionados <- gem_especialistas %>%
  filter(economy %in% países) %>%
  select(economy, year, gov_support, taxes_bureaucracy,
         gov_programs, internal_market_dynamics)
```

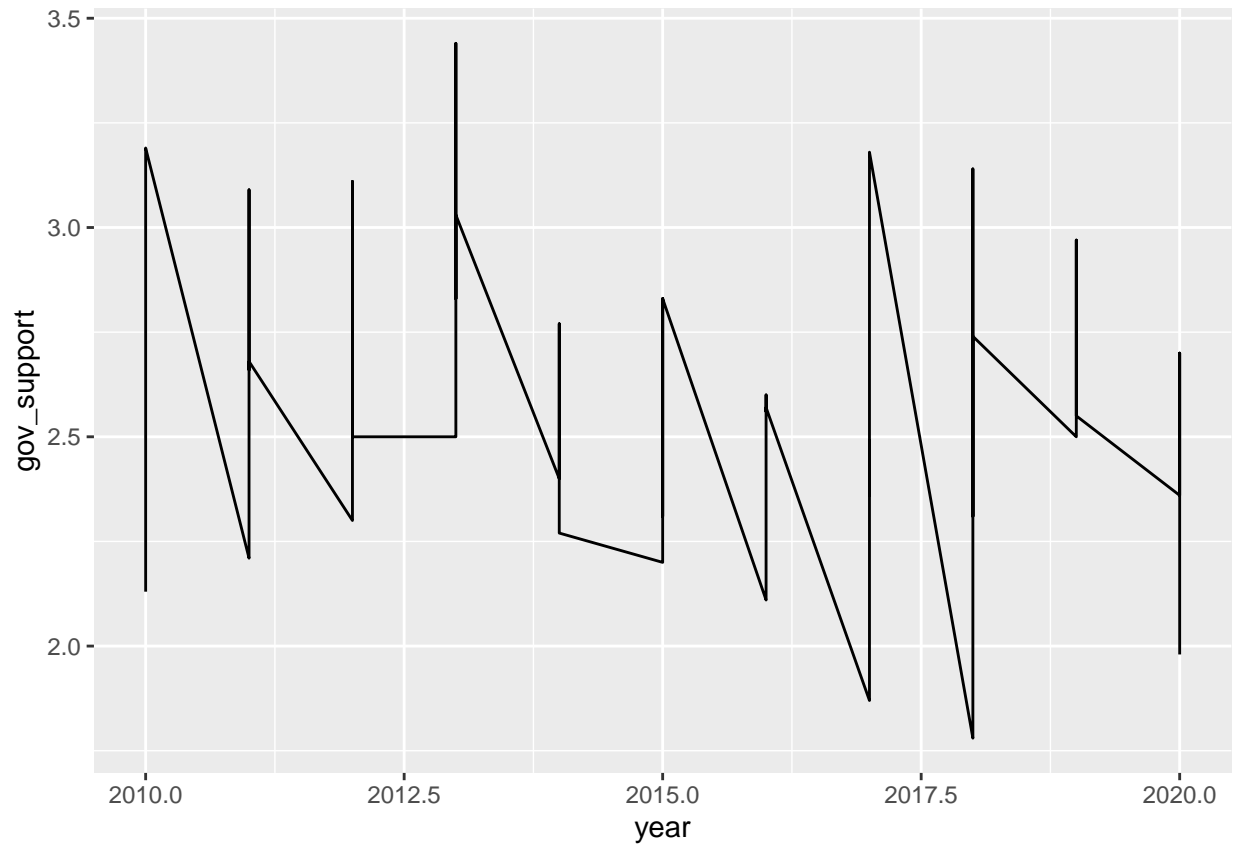
## Estatísticas descritivas

```
gem_países_selecionados %>%
  select(-year) %>%
  group_by(economy) %>%
  skim()
```

## Plotando variáveis

Vamos investigar a variável `gov_support` dos países ao longo do tempo.

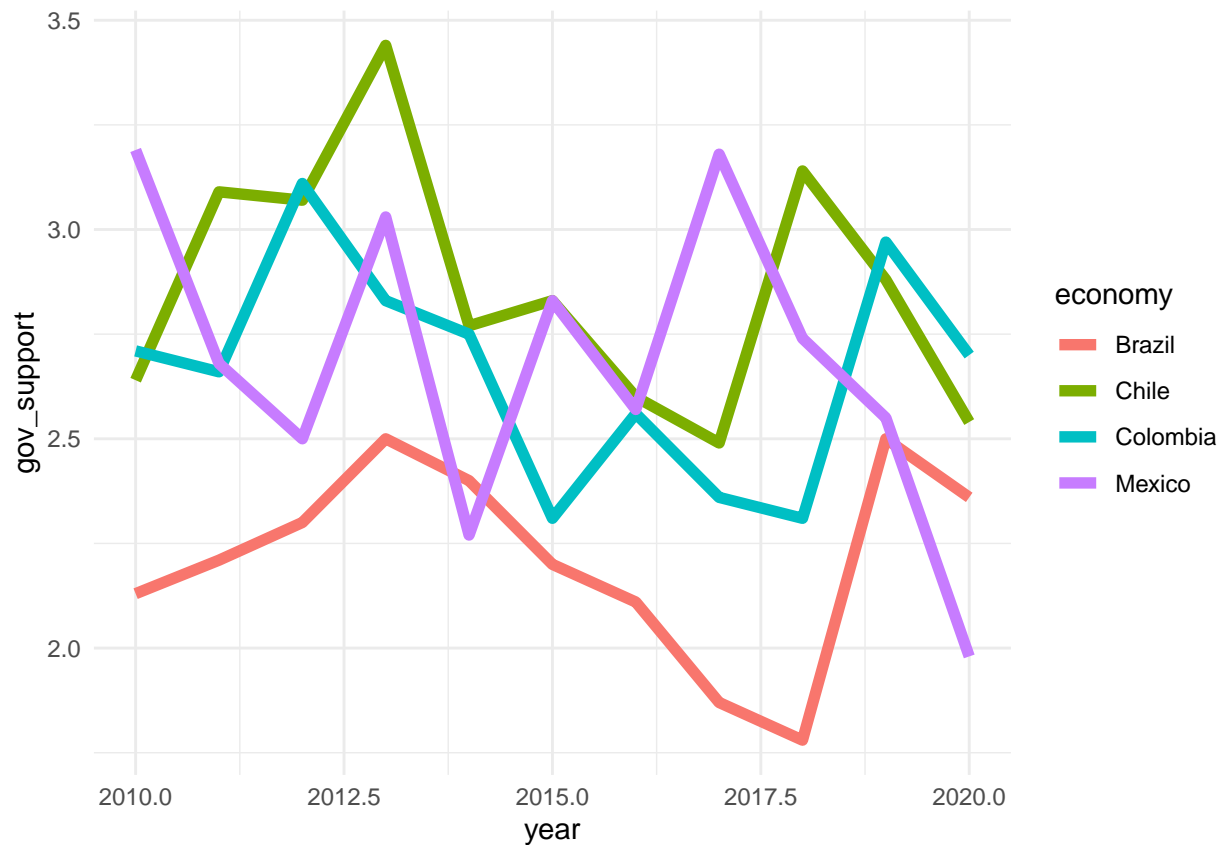
```
gem_países_selecionados %>%
  ggplot(aes(year, gov_support)) + geom_line()
```



## Adicionando o parâmetro col

Vamos ainda adicionar e destacar as linhas por país e retirar esse fundo cinza.

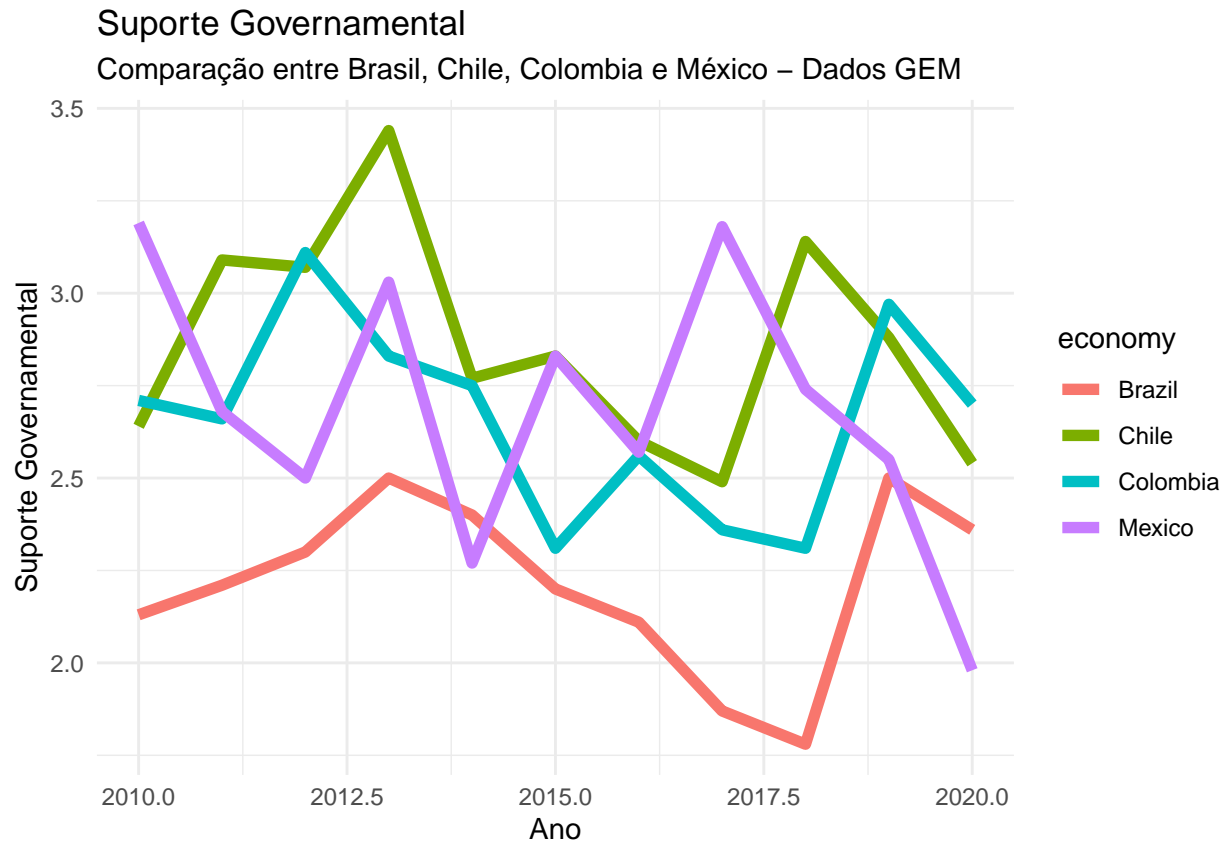
```
# a função col é usada aqui para colorir as linhas
gem_países_selecionados %>%
  ggplot(aes(year, gov_support, col = economy)) + geom_line(size = 2) +
  theme_minimal()
```



## Adicionando títulos

Para adicionar título, vamos usar o `ggtitle`, além disso, iremos usar as funções `xlab` e `ylab` para mudar os nomes dos eixos x e y.

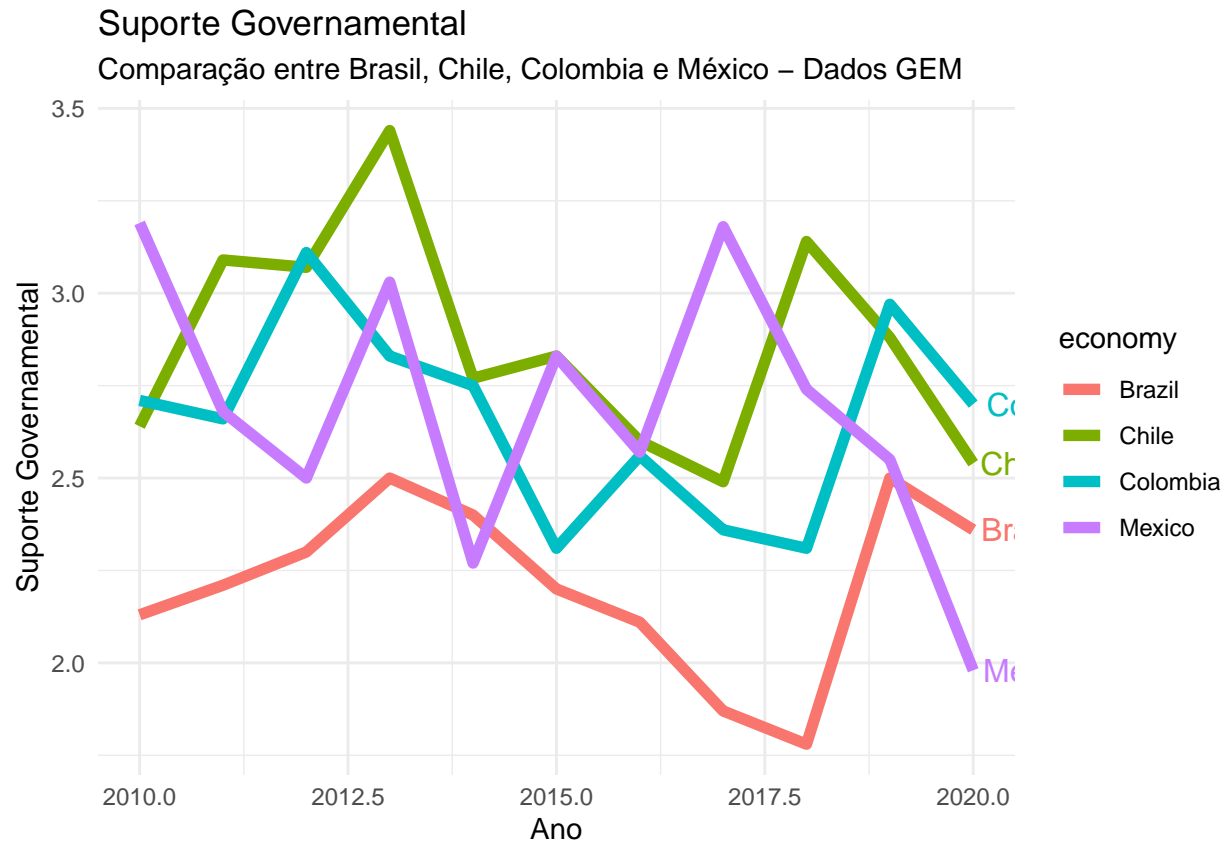
```
# Estamos definindo os rótulos do gráfico e adicionando título.
gem_paises_selecionados %>%
  ggplot(aes(year, gov_support, col = economy)) + geom_line(size = 2) +
  theme_minimal() + xlab("Ano") + ylab("Suporte Governamental") +
  ggtitle("Suporte Governamental",
          "Comparação entre Brasil, Chile, Colombia e México - Dados GEM")
```



## Adicionando textos às linhas

Vamos agora usar o pacote `directlabels` para adicionar o nome do país associado a cada linha.

```
gem_países_selecionados %>%
  ggplot(aes(year, gov_support, col = economy)) + geom_line(size = 2) +
  geom_dl(aes(label = economy), method = list(dl.combine("last.points"), hjust = -0.1)) +
  theme_minimal() + xlab("Ano") + ylab("Suporte Governamental") +
  ggtitle("Suporte Governamental",
    "Comparação entre Brasil, Chile, Colombia e México - Dados GEM")
```



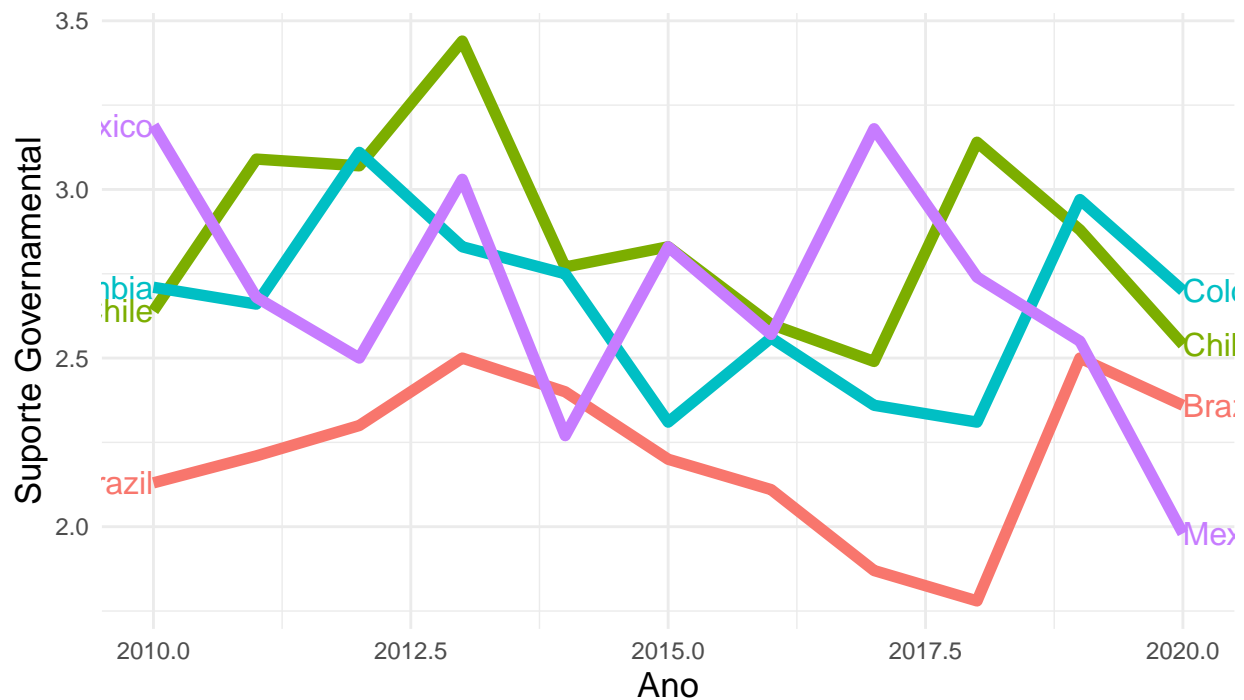
## Aumentando fontes

Vamos aumentar as fontes e retirar a legenda usando a função `theme`.

```
# Estamos usando a funcao theme para modificar alguns elementos do grafico.
# Neste caso, o tamanho da fonte e retirando a legenda.
gem_paises_selecionados %>%
  ggplot(aes(year, gov_support, col = economy)) + geom_line(size = 2) +
  geom_dl(aes(label = economy), method = list(dl.combine("first.points", "last.points")) +
  theme_minimal() + xlab("Ano") + ylab("Suporte Governamental") +
  theme(legend.position = "none") +
  theme(plot.title = element_text(size=22)) +
  theme(axis.title.y = element_text(size = rel(1.2), angle = 90)) +
  theme(axis.title.x = element_text(size = rel(1.2), angle = 0)) +
  ggtitle("Suporte Governamental", "Comparação entre Brasil, Chile,
  Colombia e México - Dados GEM")
```

# Suporte Governamental

Comparação entre Brasil, Chile, Colombia e México – Dados GEM



## Mudando a escala

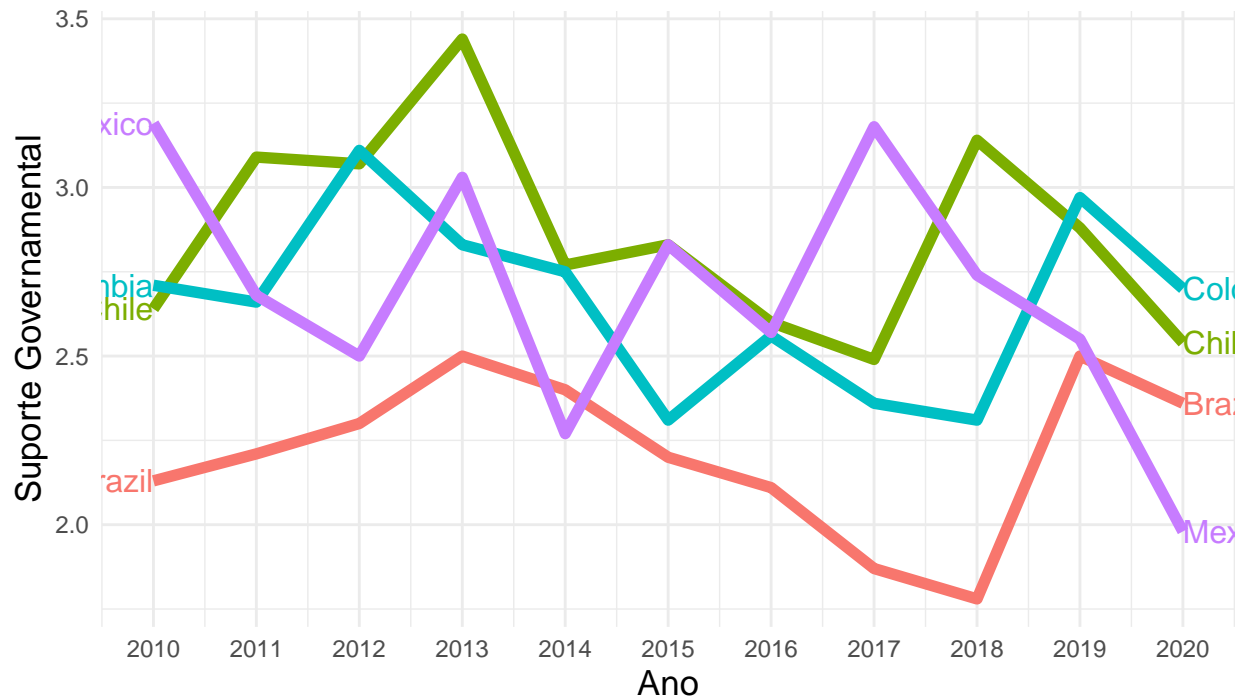
Além de aumentar as fontes e retirar a legenda, vamos também usar o argumento `scale_x_continuous` para mudar a escala com o parâmetro `breaks`.

```
gem_paises_selecionados %>%  
  ggplot(aes(year, gov_support, col = economy)) + geom_line(size = 2) +  
  geom_dl(aes(label = economy), method = list(dl.combine("first.points", "last.points"))) +  
  theme_minimal() + xlab("Ano") + ylab("Suporte Governamental") +  
  theme(legend.position = "none") +  
  theme(plot.title = element_text(size=22)) +  
  theme(axis.title.y = element_text(size = rel(1.2), angle = 90)) +  
  theme(axis.title.x = element_text(size = rel(1.2), angle = 0)) +  
  ggtitle("Suporte Governamental", "Comparação entre Brasil, Chile,  
          Colombia e México - Dados GEM") +  
  scale_x_continuous(breaks = seq(2010, 2021, 1))
```



# Suporte Governamental

Comparação entre Brasil, Chile, Colombia e México – Dados GEM



## Explorando dados de Adult Population Survey

Os dados consolidados do levantamento com população adulta é composto um conjunto amplo de variáveis. Para o presente exercício vamos usar as seguintes variáveis:

- **Taxa de Oportunidades Percebidas:** Percentual de indivíduos de 18 a 64 anos (exceto aqueles envolvidos em algum estágio de atividade empreendedora) que enxergam boas oportunidades para se iniciar um negócio na área onde vivem.
- **Taxa de atividade empreendedora total (TAE):** Percentual da população de 18 a 64 anos que é um empreendedor nascente ou proprietário-gerente de um novo negócio.
- **Medo de fracassar:** Percentual da população de 18 a 64 anos (indivíduos envolvidos em qualquer estágio da atividade empresarial excluídos) que indicam que o medo do fracasso os impedem de abrir um negócio.
- **Empreendedorismo como boa escolha de carreira:** Percentual da população de 18 a 64 anos que concorda com a afirmação de que, em seu país, a maioria das pessoas considera iniciar um negócio como uma escolha de carreira desejável.

## Associando o GEM a outras bases

Muitas pesquisas utilizam o GEM associado a outras bases. Para o nosso exercício, vamos utilizar dados da Worldwide Governance Indicator (WGI) que pode ser acessada por meio da seguinte página. Antes de qualquer coisa, vamos ver a definição de governança conforme o WGI.

*“Governança consiste na tradição e nas instituições pelas quais a autoridade em um país é exercida. Isso inclui o processo pelo qual um governante é escolhido, monitorado e substituído; a capacidade do governo formular efetivamente e implementar políticas sólidas; e o respeito dos cidadãos e do Estado pelas instituições que regem as interações econômicas e sociais entre eles.” (WGI) As variáveis variam entre -2,5 (fraco índice de governança) a 2,5 (forte índice de governança). Para fins didáticos, transformei o indicador para variar de 0 a 5 pontos.*

## Conhecendo variáveis do WGI

- **Control of Corruption:** Reflete as percepções da extensão em que o poder público é exercido para ganho privado, incluindo formas insignificantes e grandes de corrupção, bem como a “captura” do Estado pelas elites e interesses privados.
- **Rule of Law:** Reflete as percepções de até que ponto os agentes confiam e cumprem as regras da sociedade e, em particular, a qualidade da execução de contratos, direitos de propriedade, polícia e tribunais, bem como a probabilidade de crime e violência.
- **Regulatory Quality:** Reflete a percepção da capacidade do governo de formular e implementar políticas e regulamentações sólidas que permitam e promovam o desenvolvimento do setor privado.
- **Political Stability and Absence of Violence/Terrorism:** A Estabilidade Política e Ausência de Violência/Terrorismo mede as percepções da probabilidade de instabilidade política e/ou violência com motivação política, incluindo terrorismo.
- **Voice and Accountability:** Reflete as percepções de até que ponto os cidadãos de um país podem participar da seleção de seu governo, bem como a liberdade de expressão, a liberdade de associação e a liberdade de mídia.

## Juntando bases

### Lendo as bases do WGI e do GEM a partir do GitHub

```
#Base do WGI:
wgidataset <- read_csv("https://raw.githubusercontent.com/danielppagotto/R_empreendedorismo2/main/arqui

## Warning: Missing column names filled in: 'X1' [1]

##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   country = col_character(),
##   code = col_character(),
##   corruption = col_double(),
##   rule_of_law = col_double(),
##   regulatory_quality = col_double(),
##   gov_effectiveness = col_double(),
##   political_stability = col_double(),
##   voice_accountability = col_double()
## )
```

*#Base do GEM:*

```
gem_aps <- read_delim("https://raw.githubusercontent.com/danielppagotto/R_empreendedorismo2/main/arquivo")
```

```
##
## -- Column specification -----
## cols(
##   id = col_double(),
##   cod_pais = col_double(),
##   economy = col_character(),
##   continent = col_character(),
##   abbrev = col_character(),
##   year = col_double(),
##   perceived_opportunities = col_double(),
##   perceived_capabilities = col_double(),
##   fear_failure = col_double(),
##   entrepreneurial_intentions = col_double(),
##   tea = col_double(),
##   established_ownership = col_double(),
##   entrepreneurial_employee_Activity = col_double(),
##   female_male_tea = col_double(),
##   high_job_creation_expectation = col_double(),
##   business_service_sector = col_double(),
##   high_status_successful_entrepreneurs = col_double(),
##   entrepreneurship_as_good_carrer_choice = col_double()
## )
```

Com este comando iremos selecionar as variáveis de interesse.

```
gem_wgid <- gem_aps %>%
  left_join(wgidataset, by = c("abbrev" = "code")) %>%
  select(economy, continent, tea, perceived_opportunities,
         fear_failure, established_ownership,
         entrepreneurship_as_good_carrer_choice,
         corruption, rule_of_law, regulatory_quality,
         political_stability, voice_accountability)
```

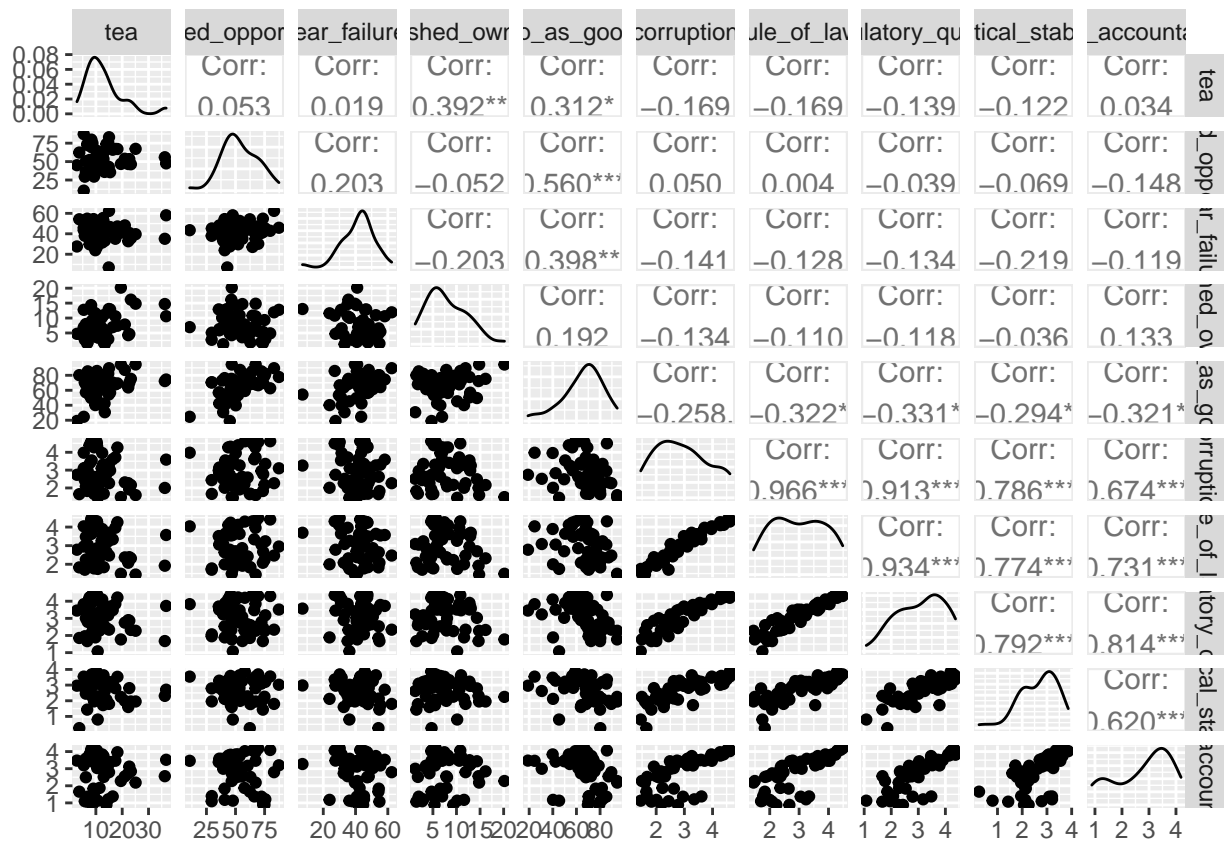
E finalmente, vamos inspecionar alguns dados de estatística descritiva por continente.

```
gem_wgid %>%
  group_by(continent) %>%
  skim()
```

## Estatística descritiva - GGally

Nessa etapa vamos usar a função do pacote GGally chamada ggpairs.

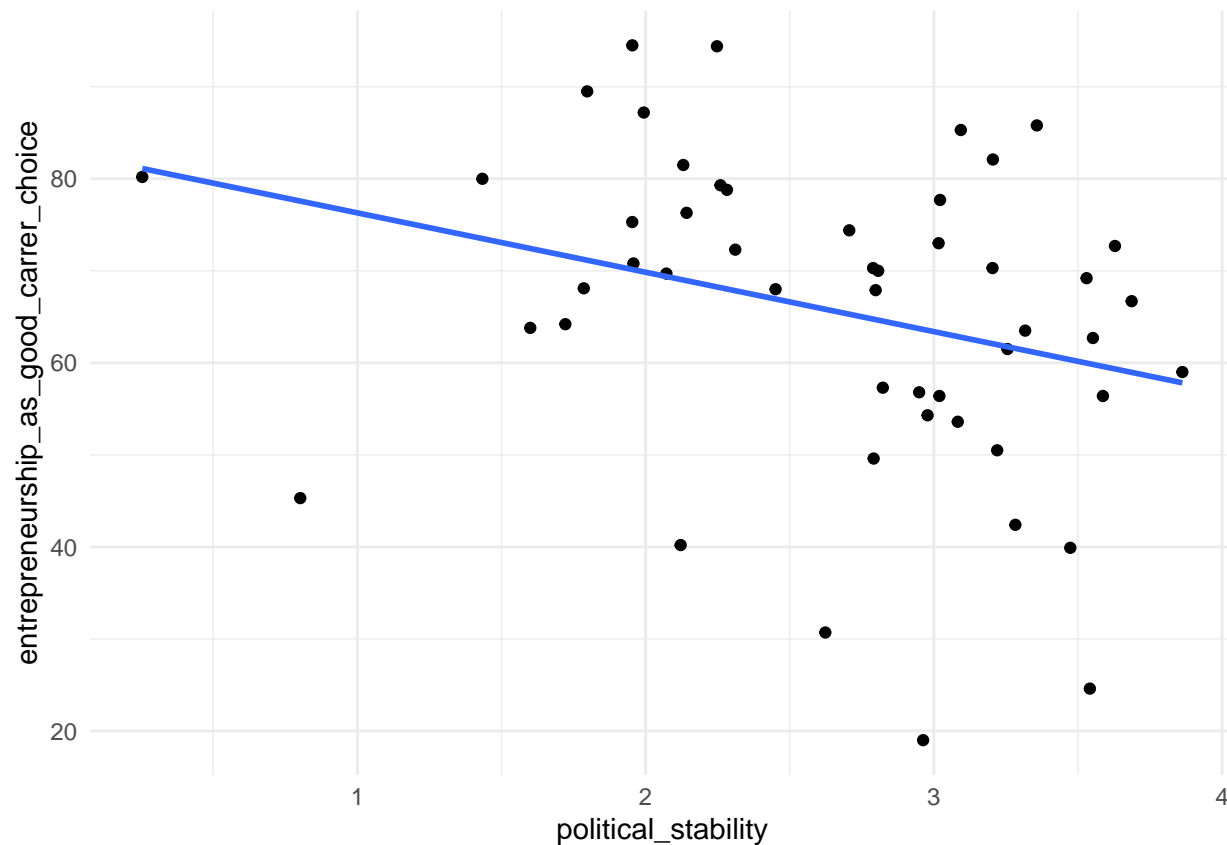
```
gem_wgid %>% select(-economy, -continent) %>%
  ggpairs()
```



Podemos ir um pouco além e explorar mais, por exemplo, a variável Entrepreneurship as good career choice.

```
gem_wgid %>%
  ggplot(aes(x = political_stability, y = entrepreneurship_as_good_carrer_choice)) +
  geom_point() + theme_minimal() + geom_smooth(method = "lm", se = FALSE)

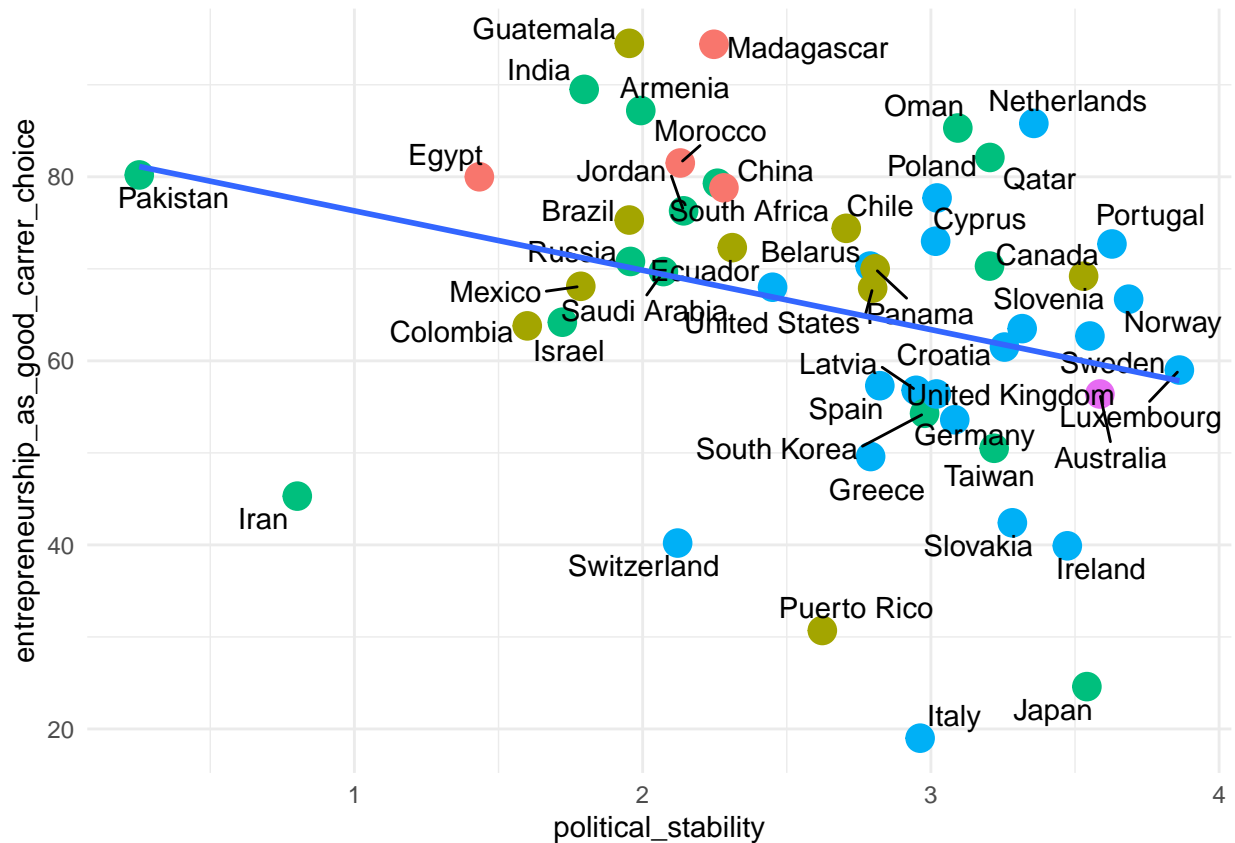
## `geom_smooth()` using formula 'y ~ x'
```



```
gem_wgid %>%
  ggplot(aes(x = political_stability, y = entrepreneurship_as_good_carrer_choice)) +
  geom_point(aes(col = continent, size = 1.5)) +
  theme_minimal() + geom_text_repel(aes(label = economy)) +
  theme(legend.position = "none") + geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

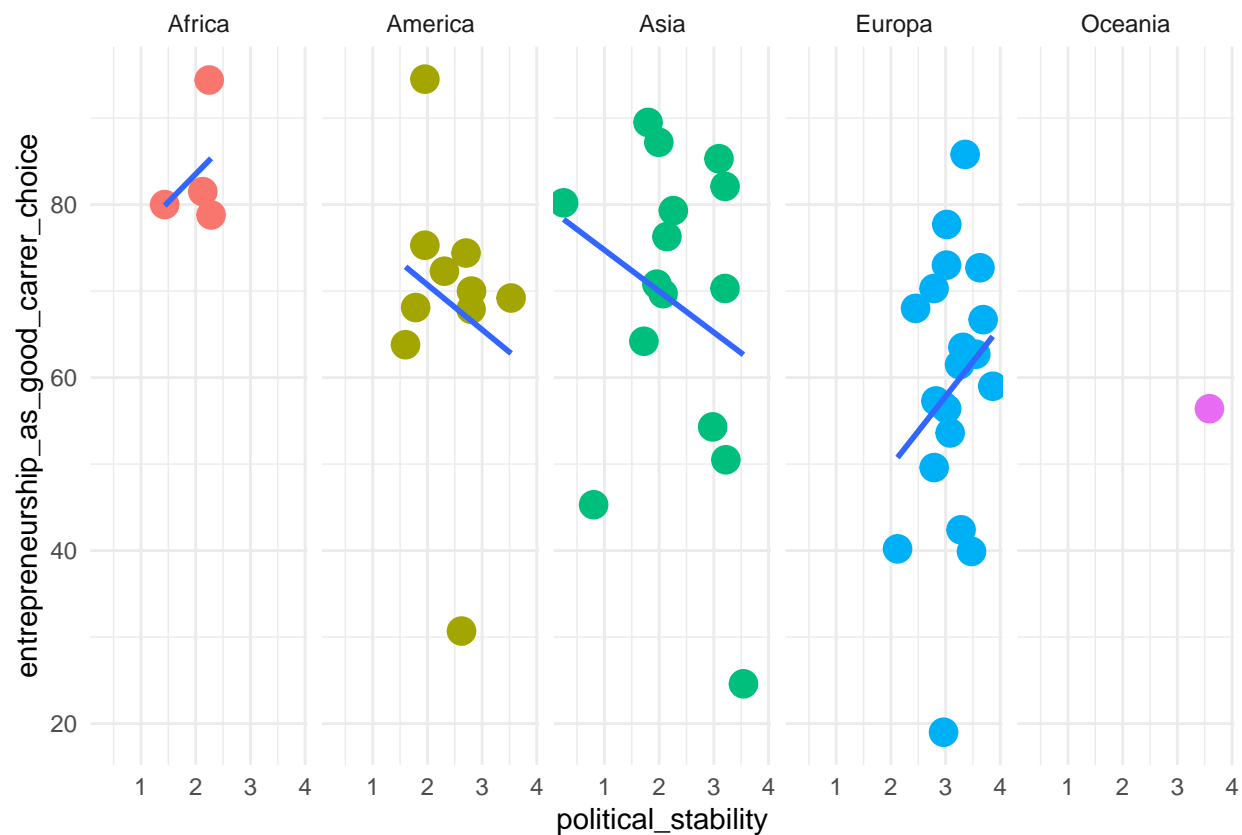


## Aplicando facet\_grid()

Não para por aí, agora vamos usar o `facet_grid` para formar uma matriz de painéis.

```
gem_wgid %>%
  ggplot(aes(x = political_stability, y = entrepreneurship_as_good_carrer_choice)) +
  geom_point(aes(col = continent, size = 1.5)) + facet_grid(~continent) +
  theme_minimal() + theme(legend.position = "none") + geom_smooth(method = "lm", se = FALSE)
```

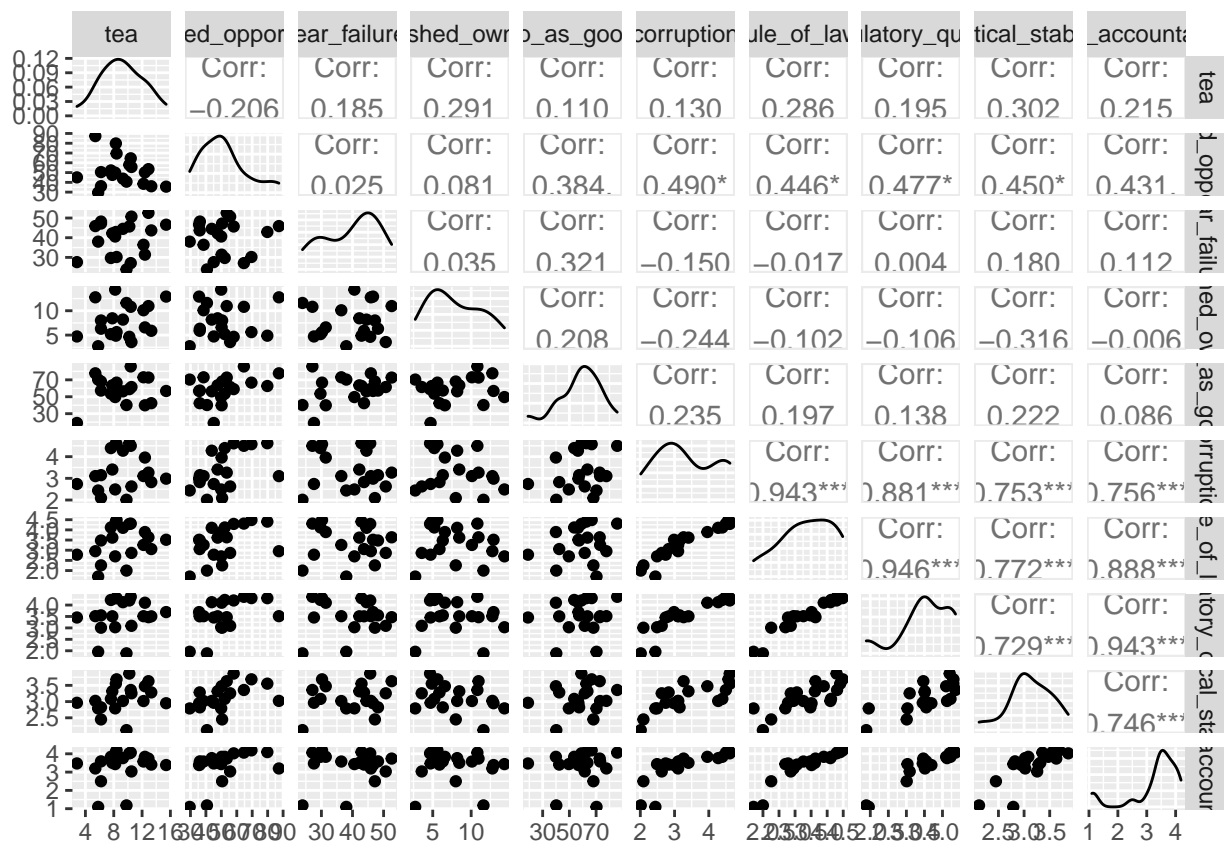
```
## `geom_smooth()` using formula 'y ~ x'
```



## Podemos voltar ao GGally

Podemos filtrar um continente apenas e verificar o resultados das variáveis.

```
#vamos pegar como exemplo a Europa.
gem_wgid %>%
  filter(continent == "Europa") %>%
  select(-economy, -continent) %>%
  ggpairs()
```



Obrigado

Se encontrou algum erro ou tem alguma sugestão de melhoria pode entrar em contato. Se foi útil também avise! Ah... e pode passar adiante. Quanto mais gente tendo acesso a esse conhecimento melhor!

Daniel Pagotto | danielppagotto@gmail.com | LinkedIn

LAPEI - UFG | lapeiufg@gmail.com. | No instagram: lapeiufg