VISUALIZAÇÃO DE DADOS APLICADA

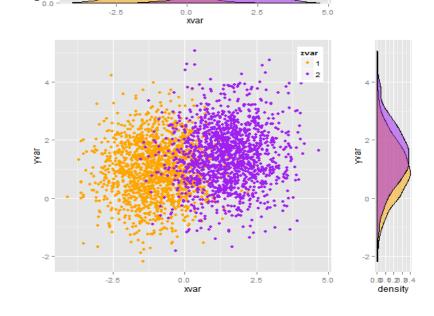


Prof. Dr. Anderson Ara

Slide 07

O pacote ggplot2, criado por Hadley Wickham, oferece uma poderosa linguagem gráfica para criar gráficos elegantes e complexos. Sua popularidade na comunidade R explodiu nos últimos anos e é – **hoje** - a principal implementação da gramática de gráficos para o R. Geralmente baseado na Gramática de Gráficos de Leland Wilkinson.

O ggplot2 permite criar gráficos que representam dados numéricos e categóricos univariados e multivariados de forma direta. O agrupamento pode ser representado por cor, símbolo, tamanho e transparência. A criação de parcelas de trellis (isto é, condicionamento) é também é relativamente simples.



Fornece uma estrutura unificadora (uma gramática) para descrever e especificar gráficos

Cria gráficos de dados de forma incremental a partir de pequenos pedaços de código.

NA GRAMÁTICA GGPLOT2:

Uma aesthetic (estética) é um mapeamento explícito entre uma variável e os elementos visuais que representam seus valores.

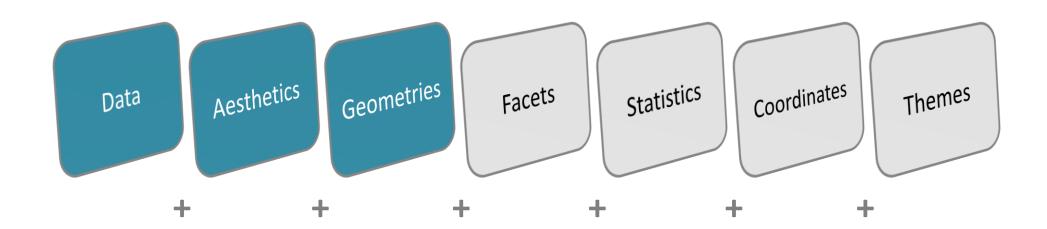
Um glyph (glifo) é o elemento gráfico básico que representa um caso (outros termos utilizados incluem "mark" e " symbol").

Em um gráfico de dispersão, as posições de um glifo na janela gráfica (nos sentidos horizontais e verticais) são as elementos visuais que ajudam o usuário a entender quão grandes são as quantidades correspondentes.

A estética é o mapeamento que descreve essas correspondências. Quando mais de duas variáveis estão presentes, a estética adicional pode gerar sugestões visuais adicionais. Observe também que algumas elementos visuais (como a direção em uma série temporal) estão implícitos e não possuem uma estética correspondente.

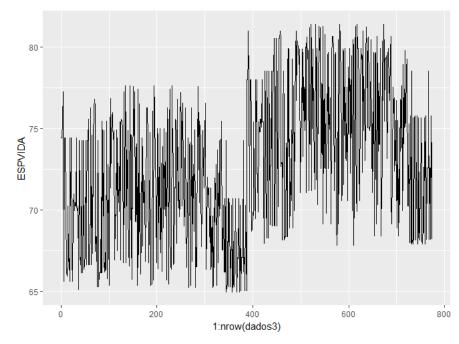
GRAMÁTICA GGPLOT2

INSTRUÇÃO EM CAMADAS:

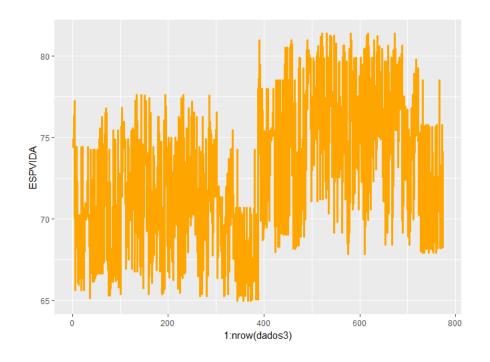


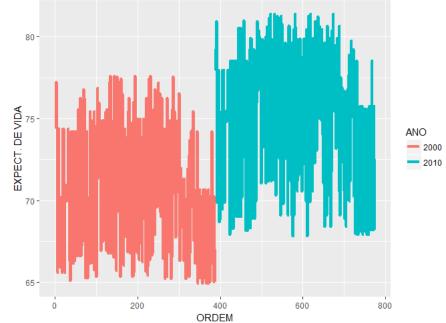
```
#PRIMEIROS PASSOS
require(ggplot2)
head(dados3)

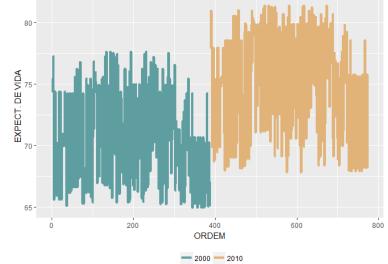
ggplot(data = dados3, aes(y = ESPVIDA, x = 1:nrow(dados3))) +
geom_line()
```



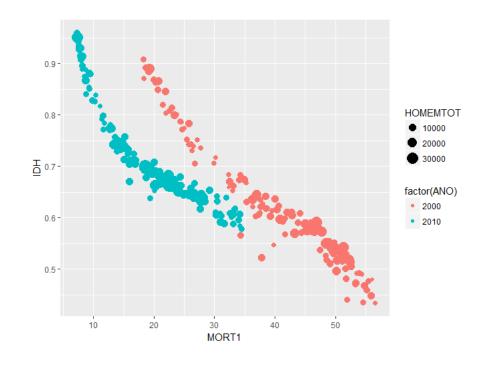
```
g1 <-ggplot(data = dados3, aes(y = ESPVIDA, x =
1:nrow(dados3))) + geom_line()
g1 <- g1 + geom_line(size=1.1,color="orange")
g1</pre>
```

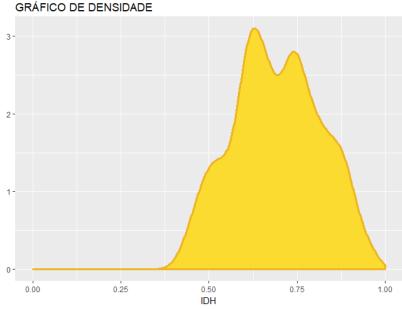




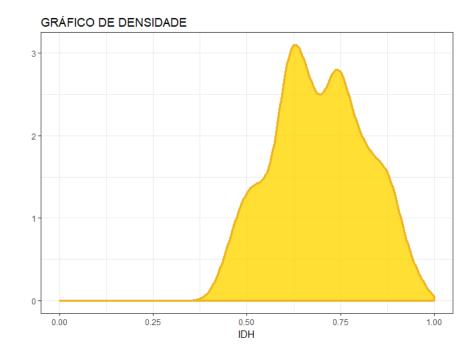


```
#SCATTERPLOT
g4 <- ggplot(data = dados3, aes(y = IDH, x = MORT1))
g4 + geom_point(size = 5)
g4 + geom_point(aes(color = factor(ANO)), size = 5)
g4 + geom_point(aes(color = factor(ANO), size = HOMEMTOT))</pre>
```

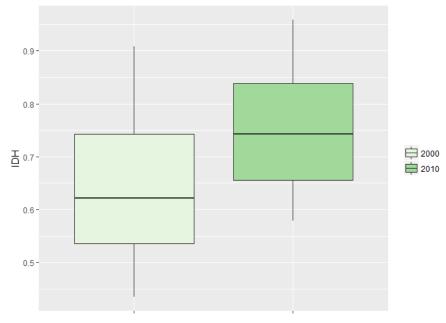




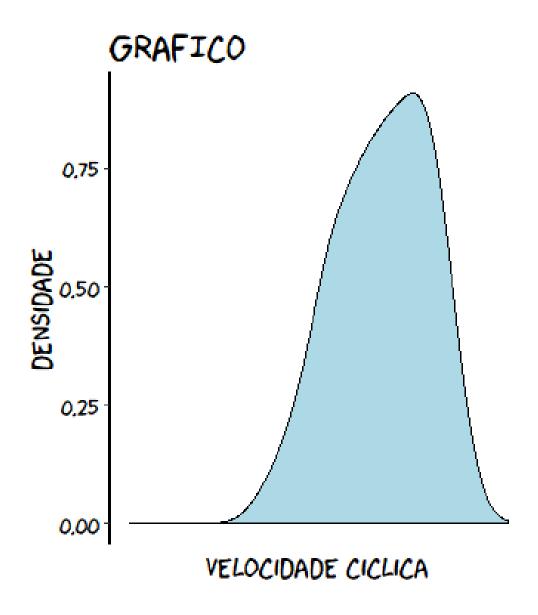
#TEMA BÁSICO



```
##BOXPLOT
gg<-
ggplot(dados3, aes(x=factor(ANO), y=IDH, fill=factor(ANO)))+geom
_boxplot()
gg<- gg + scale_x_discrete(name = "")
gg<- gg + theme(axis.text.x = element_blank())
gg<- gg + scale_fill_brewer(palette="Greens")
gg<- gg + theme(legend.title = element_blank())
gg</pre>
```

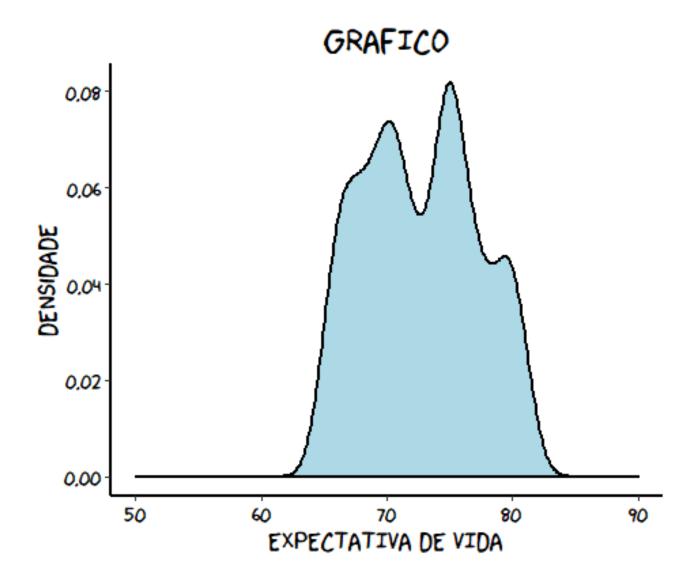


DENSIDADE



```
library(extrafont)
##DOWNLOAD DA FONTE
download.file("http://simonsoftware.se/other/xkcd.
ttf", dest="xkcd.ttf", mode="wb")
system("cmd.exe", input = "mkdir C:\\fonts")
system("cmd.exe", input = "copy xkcd.ttf
C:\\fonts")
font import(paths = "C:\\fonts",
pattern="[X/x]kcd")
fonts()
loadfonts()
```

```
##CRIANDO NOVO GRÁFICO COM A FONTE
qq2 \leftarrow qqplot(dados3, aes(x = ESPVIDA)) +
  geom density(colour = "black", fill = "lightblue", size=1) +
  scale x continuous (name = "EXPECTATIVA DE VIDA",
                     limits=c(50, 90)) +
  scale y continuous(name = "Densidade") +
  ggtitle("GRAFICO") +
  theme(axis.line = element line(size=1, colour = "black"),
        panel.grid.major = element blank(),
        panel.grid.minor = element blank(),
        panel.border = element blank(),
        panel.background = element blank(),
        plot.title=element text(size = 20, family="xkcd", hjust
= 0.5),
        text=element text(size = 16, family="xkcd"),
        axis.text.x=element text(colour="black", size = 12),
        axis.text.y=element text(colour="black", size = 12))
gg2
```



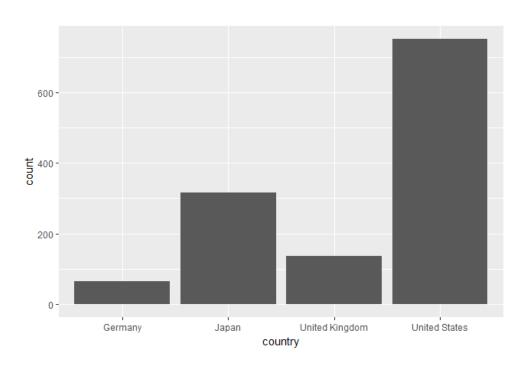
#PREPARANDO OS DADOS

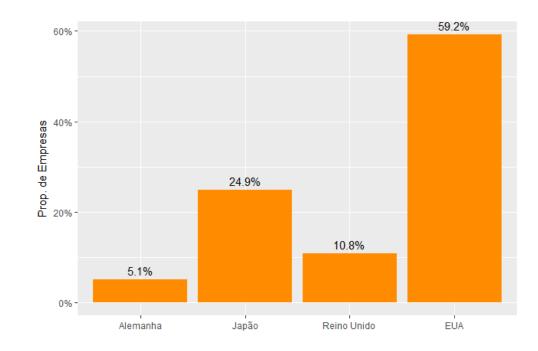
```
require(ggplot2)
require(HSAUR)
data(Forbes2000)
sel=Forbes2000[,"country"] == "Germany" |
  Forbes2000[, "country"] == "United Kingdom"
  Forbes2000[,"country"] == "United States" |
  Forbes2000[,"country"]=="Japan"
pontos=c(min(Forbes2000$marketvalue) -
0.01, median (Forbes2000$marketvalue), max(Forbes2000$marketvalue))
Forbes2000$mk=cut(Forbes2000$marketvalue,pontos)
dados=Forbes2000[sel,]
```

GRÁFICO DE BARRAS

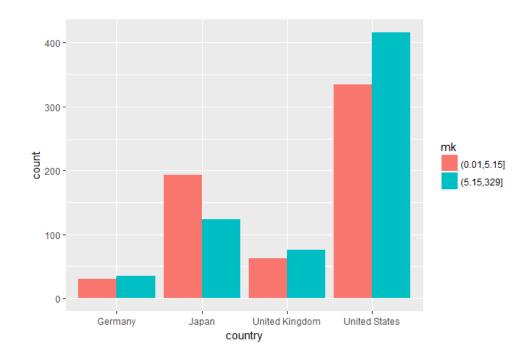
#GRÁFICO DE BARRAS

ggplot(dados) +
 aes(x=country) +
 geom_bar()



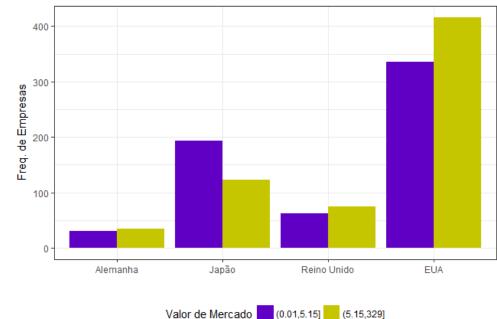


#GRÁFICO FINAL



#GRÁFICO DE BARRAS CONJUNTAS

ggplot(dados) +
 aes(x=country,fill=mk) +
 geom_bar(position="dodge")



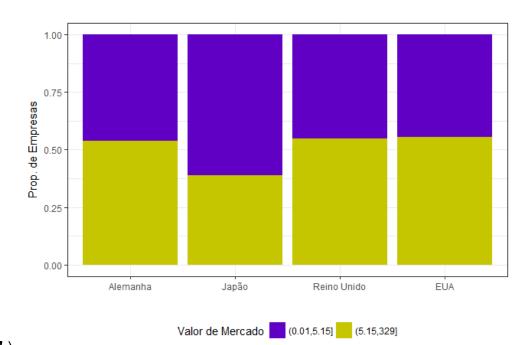
```
#GRÁFICO EDITADO
```

```
cores <- c("#6000C5", "#C5C500")
ggplot(dados)+
  aes(x=country,fill=mk) +
   geom_bar(position="dodge") +
   xlab(" ") +
   ylab("Freq. de Empresas")+
   scale_x_discrete(labels=c("Alemanha","Japão","Reino Unido","EUA")) +
   scale_fill_manual("Valor de Mercado", values = cores)+
   theme_bw()+
   theme(legend.position="bottom", legend.direction="horizontal")</pre>
```

0.75 - 0.50 - 0.50 - 0.00 - Germany Japan United Kingdom United States

#GRÁFICO DE BARRAS ACUMULADAS

cores <- c("#6000C5", "#C5C500")
ggplot(dados) +
 aes(x=country, fill=mk) +
 geom_bar(position="fill")</pre>



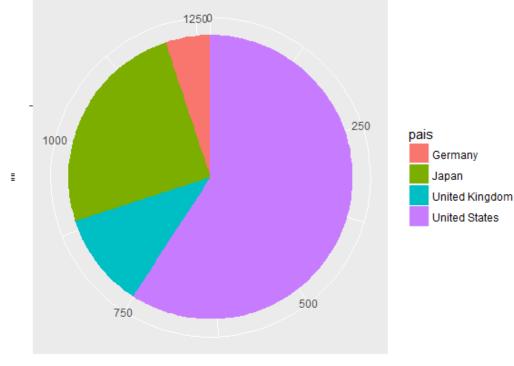
```
cores <- c("#6000C5", "#C5C500")
ggplot(dados)+
  aes(x=country,fill=mk) +
   geom_bar(position="fill") +
   xlab(" ") +
   ylab("Prop. de Empresas")+
   scale_x_discrete(labels=c("Alemanha", "Japão", "Reino Unido", "EUA")) +
   scale fill manual("Valor de Mercado", values = cores)+</pre>
```

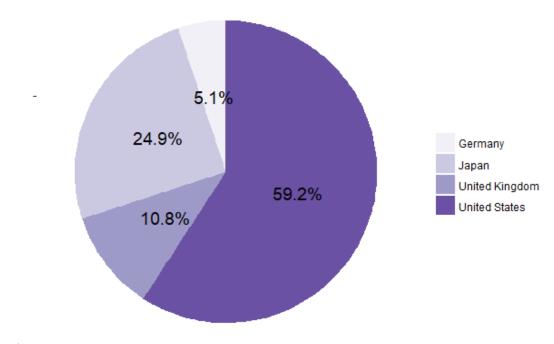
theme(legend.position="bottom", legend.direction="horizontal")

#GRÁFICO EDITADO

theme bw()+

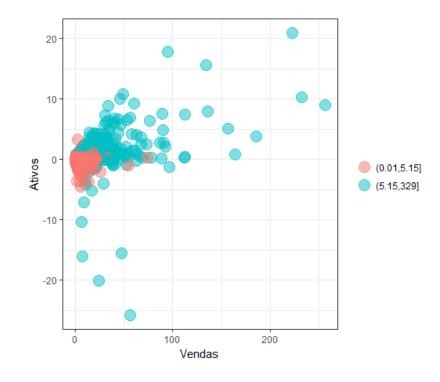
GRÁFICO DE SETORES





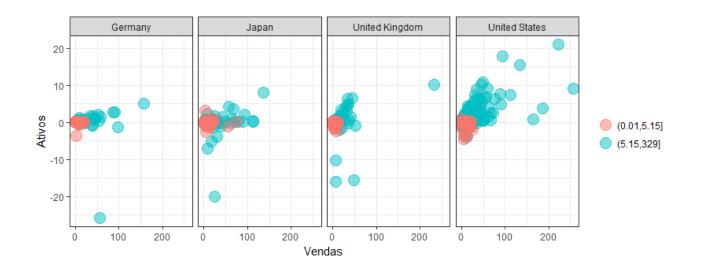
#GRÁFICO EDITADO

TRELIÇAS

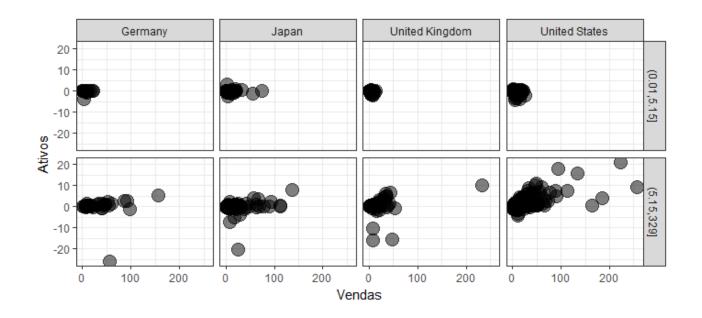


#SCATTERPLOT SIMPLES

```
ggplot(data = dados)+
  aes(x = sales, y = profits, colour=mk) +
  xlab("Vendas")+
  ylab("Ativos")+
  geom_point(size = 5, alpha=0.5)+
  theme_bw()+
  theme(legend.title = element blank())
```

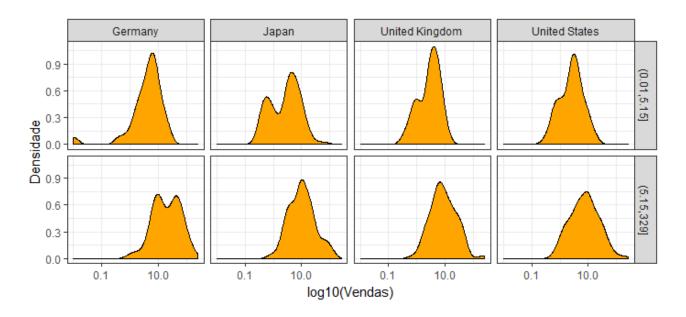


```
#TRELIÇA SIMPLES PARA SCATTERPLOT
ggplot(data = dados)+
  aes(x = sales, y = profits, colour=mk) +
  xlab("Vendas")+
  ylab("Ativos")+
  geom_point(size = 5, alpha=0.5)+
  facet_grid(facets=. ~ country)+
  theme_bw()+
  theme(legend.title = element_blank())
```



#TRELIÇA DUPLA PARA SCATTERPLOT

```
ggplot(data = dados)+
  aes(x = sales, y = profits) +
  xlab("Vendas")+
  ylab("Ativos")+
  geom_point(size = 5,alpha=0.5)+
  facet_grid(facets=mk~ country)+
  theme_bw()
```

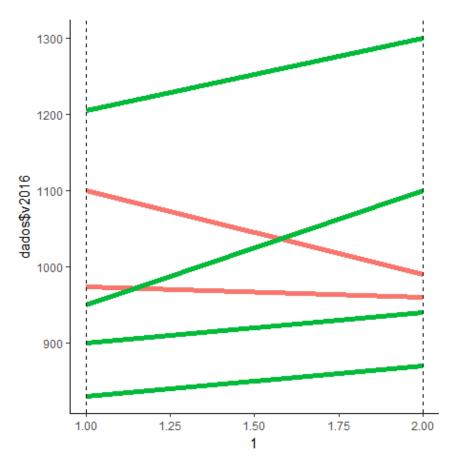


#TRELIÇA DUPLA PARA DENSIDADE

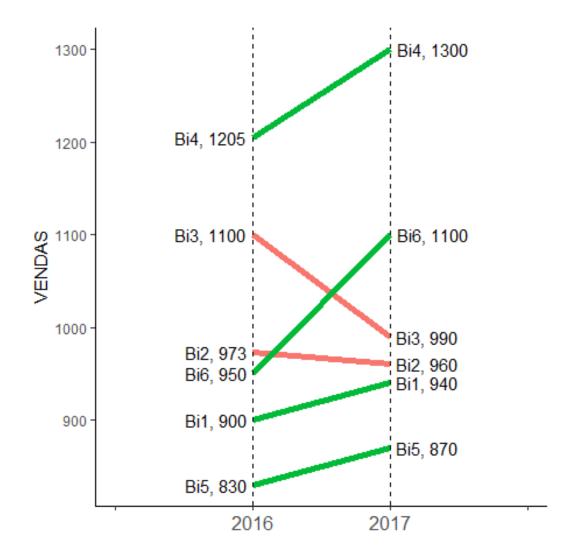
```
ggplot(data = dados) +
  aes(x = sales) +
  geom_density(fill="orange") +
  xlab("log10(Vendas)") +
  ylab("Densidade") +
  scale_x_log10() +
  facet_grid(facets=mk~ country) +
  theme_bw()
```

SLOPE CHART

```
#CONJUNTO DE DADOS
v2016=c(900,973,1100,1205,830,950)
v2017=c (940,960,990,1300,870,1100)
m <- c('Bi1','Bi2','Bi3','Bi4','Bi5','Bi6')
dados=data.frame(m, v2016, v2017)
dados$cores <- ifelse((dados$v2017 - dados$v2016) < 0, "red", "green")
theme set(theme classic())
#GRÁFICO INICIAL
p<-qqplot(dados) +</pre>
  geom segment (aes (x=1, xend=2, y=dados$v2016, yend=v2017, col=cores),
size=2, show.legend=F) +
  geom vline(xintercept=1, linetype="dashed", size=0.5) +
  geom vline(xintercept=2, linetype="dashed", size=0.5) +
  scale color manual(values = c("#00ba38", "#f8766d"))
```



р

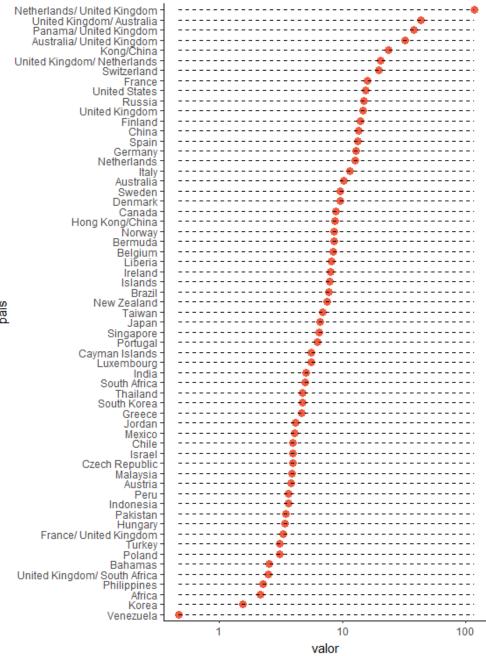


DOTPLOT

```
data(Forbes2000)
#PREPARANDO OS DADOS
valor.medio=aggregate(Forbes2000$marketvalue,
by=list(Forbes2000$country), FUN=mean)
colnames(valor.medio) = c("pais", "valor")
valor.medio=valor.medio[order(valor.medio$valor), ]
valor.medio$pais=factor(valor.medio$pais, levels = valor.medio$pais)
#GRÁFICO
ggplot(valor.medio, aes(x=pais, y=valor)) +
 geom point(col="tomato2", size=3) +
 geom segment(aes(x=pais, xend=pais, y=min(valor),
     yend=max(valor)),linetype="dashed", size=0.1) +
  scale y log10() +
 labs(title="Dot Plot",
       subtitle="País Vs Valor de Mercado Médio",
       caption="source: UFBA") +
  coord flip()
```

Dot Plot

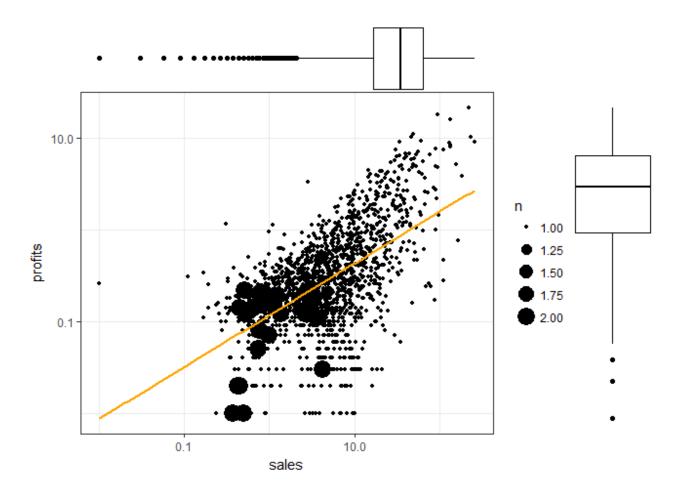


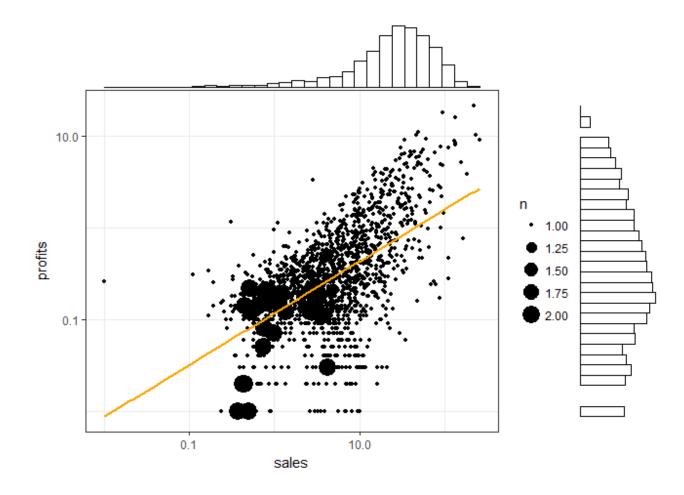


source: UFBA

GRÁFICO MARGINAIS

```
theme set(theme bw())
#PACOTE ggExtra
library(ggExtra)
#GRÁFICO
g<-ggplot(Forbes2000, aes(sales, profits)) +</pre>
  geom count() +
  scale x log10() +
  scale_y_log10()+
  geom smooth(method="lm", se=F,col="orange")
#PLOT DAS MARGINAIS
ggMarginal(g, type = "histogram", fill="transparent")
ggMarginal(g, type = "boxplot", fill="transparent")
```





BARRAS DIVERGENTES

```
theme set(theme bw())
#PREPARAÇÃO DOS DADOS
data(Forbes2000)
valor.medio=aggregate (Forbes2000$marketvalue,
by=list(Forbes2000$country), FUN=mean)
colnames(valor.medio) = c("pais", "valor")
valor.medio$valor z=(valor.medio$valor-
mean(valor.medio$valor))/sd(valor.medio$valor)
valor.medio$tipo=ifelse(valor.medio$valor z < 0, "abaixo", "acima")</pre>
valor.medio=valor.medio[order(valor.medio$valor z), ]
valor.medio$pais=factor(valor.medio$pais, levels = valor.medio$pais)
#GRÁFTCO
ggplot(valor.medio) +
  aes(x=pais, y=valor z, label=valor z, fill=tipo) +
  geom bar(stat='identity', width=.5)+
  scale fill manual (name="Valor Médio",
                    labels = c("Acima da média", "Abaixo da Média"),
                    values = c("#32CD32", "#B22222"))+
  coord flip()
```

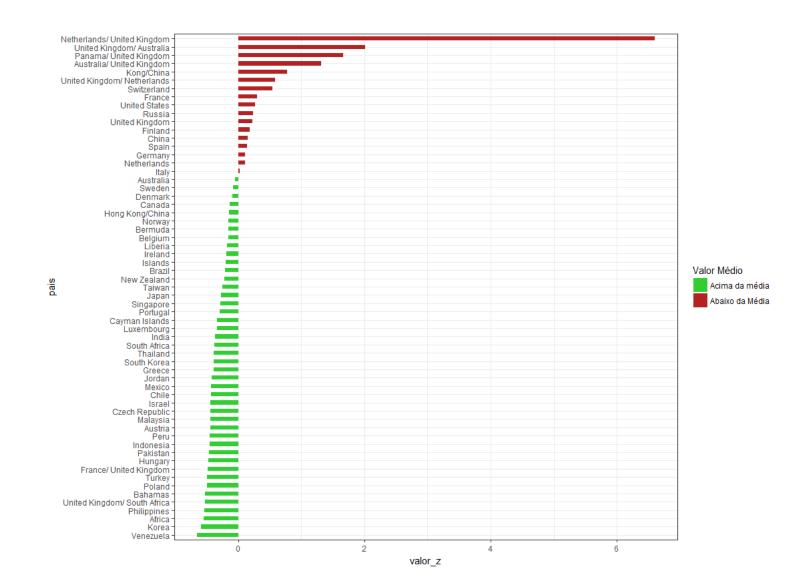
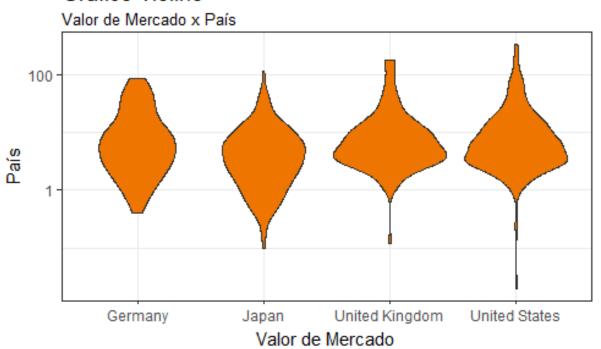


GRÁFICO DE VIOLINO

```
#PREPARANDO OS DADOS
data(Forbes2000)
sel=Forbes2000[,"country"]=="Germany"
  Forbes2000[,"country"] == "United Kingdom"
  Forbes2000[,"country"] == "United States" |
  Forbes2000[,"country"]=="Japan"
dados=Forbes2000[sel,]
theme set(theme bw())
#GRÁFICO
ggplot(dados, aes(x=country,y=marketvalue)) +
  geom violin(fill="darkorange2") +
  scale y log10() +
  labs(title="Gráfico Violino",
       subtitle="Valor de Mercado x País",
       caption="Source: UFBA",
       x="Valor de Mercado",
       y="País")
```

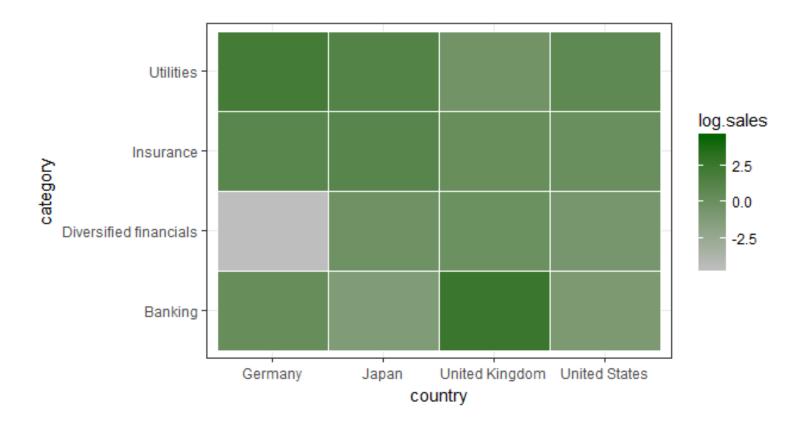
Gráfico Violino



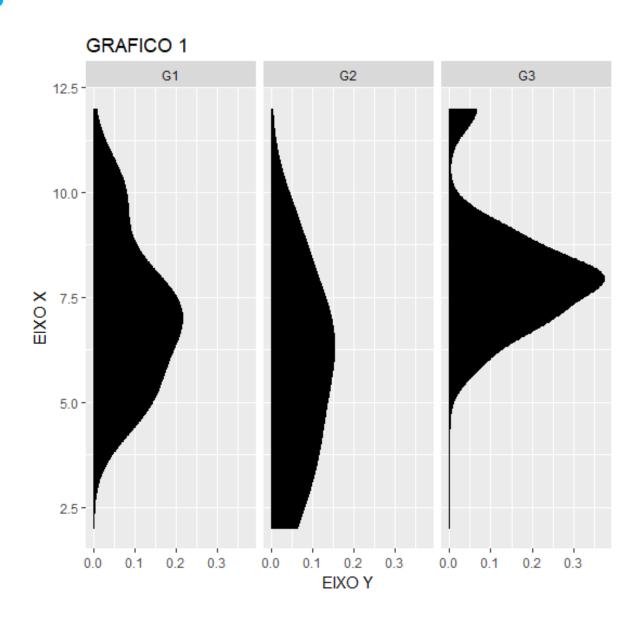
Source: UFBA

MAPA DE CALOR

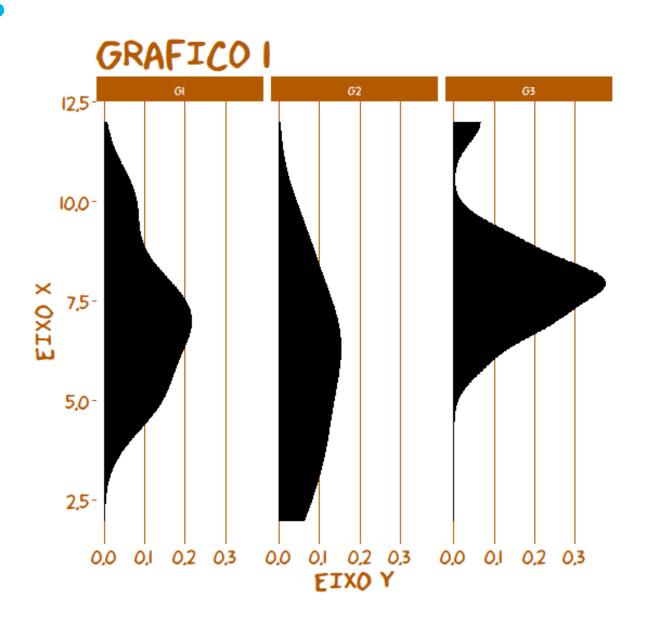
```
#PREPARANDO OS DADOS
data(Forbes2000)
sel=Forbes2000[,"country"]=="Germany" |
  Forbes2000[, "country"] == "United Kingdom" |
  Forbes2000[,"country"] == "United States" |
  Forbes2000[,"country"]=="Japan"
sel2=Forbes2000[,"category"]=="Utilities" |
  Forbes2000[,"category"] == "Diversified financials" |
  Forbes2000[, "category"] == "Insurance" |
  Forbes2000[,"category"] == "Banking"
dados=Forbes2000[sel & sel2,]
dados$log.sales=log(dados$sales)
#GRÁFICO
ggplot(dados) +
  aes(x=country, y=category) +
  geom tile(aes(fill = log.sales), colour = "white") +
  scale fill gradient(low = "gray", high = "darkgreen")
```



TEMAS



TEMAS



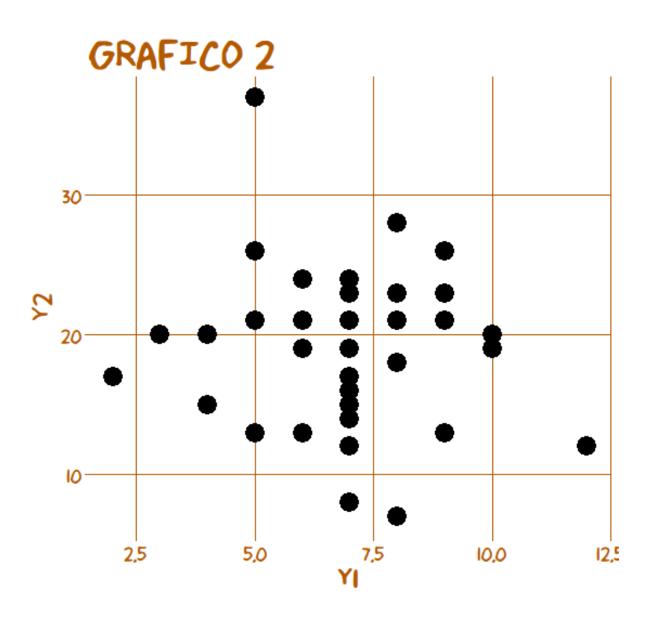
```
ggplot(data = dat)+
  aes(x = y1) +
  geom_density(fill=cor.3)+
  xlab("EIXO X")+
  ylab("EIXO Y")+
  ggtitle("GRAFICO 1")+
  facet_grid(facets=.~ grp)+
  coord_flip()
```

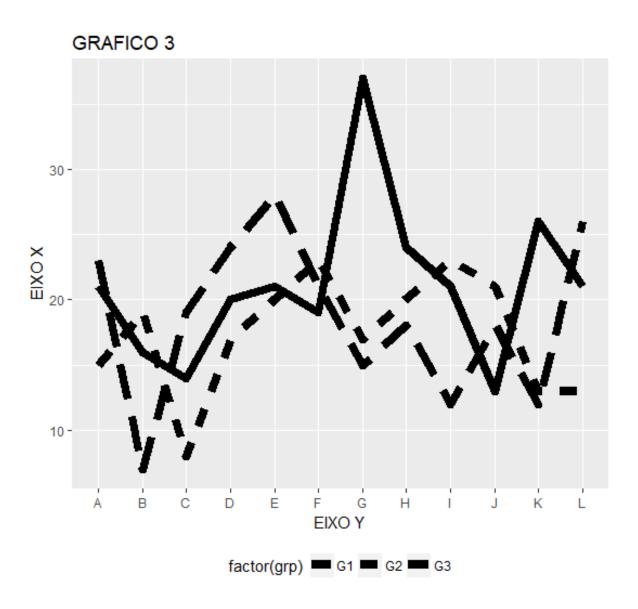
```
#LISTAGEM DE FONTES
library(extrafont)
font_import()
fonts()

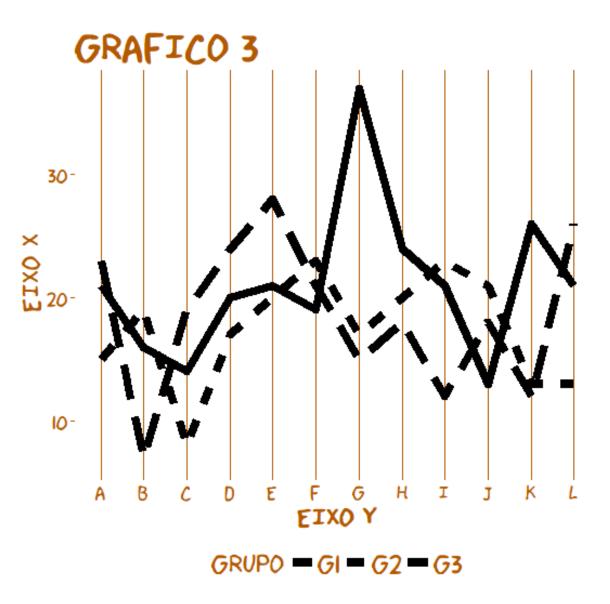
cor.1=NA
cor.2="#b35900"
cor.3="black"
```

```
tema <- function() {</pre>
 theme (
   plot.background = element rect(fill = cor.1, colour = cor.1),
   panel.background = element rect(fill = cor.1),
    axis.text = element text(colour = cor.2, family = "xkcd", size=15),
   plot.title = element text(colour = cor.2, face = "bold", size = 30,
                              vjust = 0.5, family = "xkcd"),
    axis.title = element text(colour = cor.2,
                              face = "bold", size = 20, family = "xkcd"),
   panel.grid.major.x = element line(colour = cor.2),
   panel.grid.minor.x = element blank(),
   panel.grid.major.y = element blank(),
   panel.grid.minor.y = element blank(),
    strip.text = element text(family = "xkcd", colour = "white"),
    strip.background = element rect(fill = cor.2),
    axis.ticks = element line(colour = cor.2),
    legend.title = element text(family = "xkcd", colour = cor.2,
                              size = 20).
    legend.background = element rect(fill = cor.1),
    legend.key = element rect(fill = cor.1, colour = cor.1),
    legend.text = element text(family = "xkcd", colour = cor.2, size =
20))
```

```
g1=ggplot(data = dat)+
  aes(x = y1) +
  geom_density(fill=cor.3)+
  xlab("EIXO X")+
  ylab("EIXO Y")+
  ggtitle("GRAFICO 1")+
  facet_grid(facets=.~ grp)+
  coord_flip()+
  tema()
g1
```

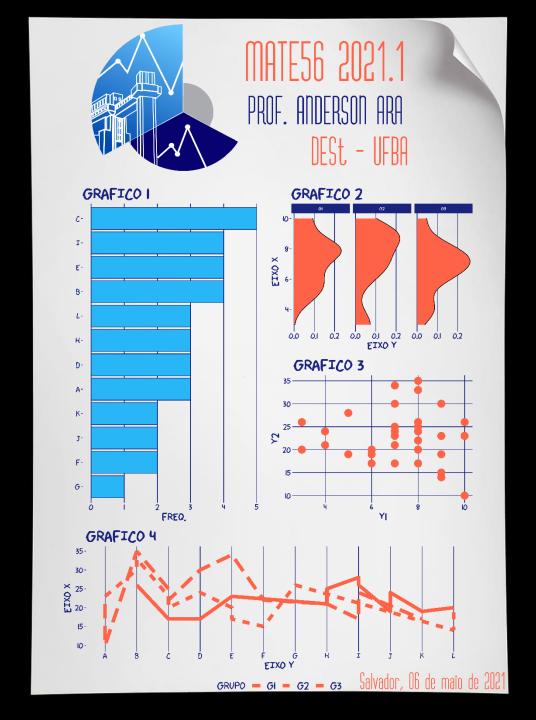






```
g2=ggplot(dat, aes(y1,y2)) +
  geom_point(size=6)+
  ggtitle("GRAFICO 2")+
  tema()+
  theme(panel.grid.major.y = element_line(colour = cor.2))
g2
```

g3

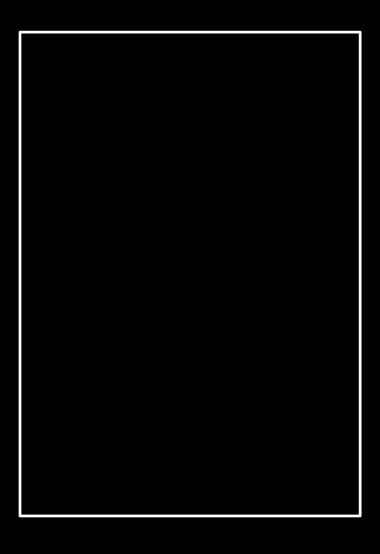


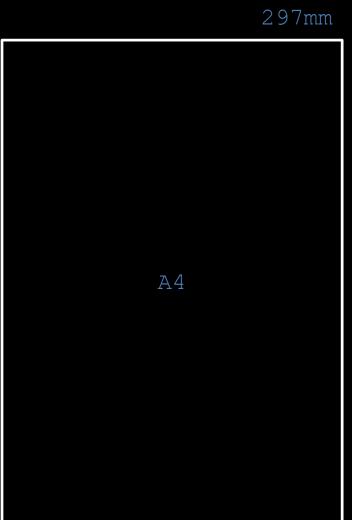


lg_estatistca.png

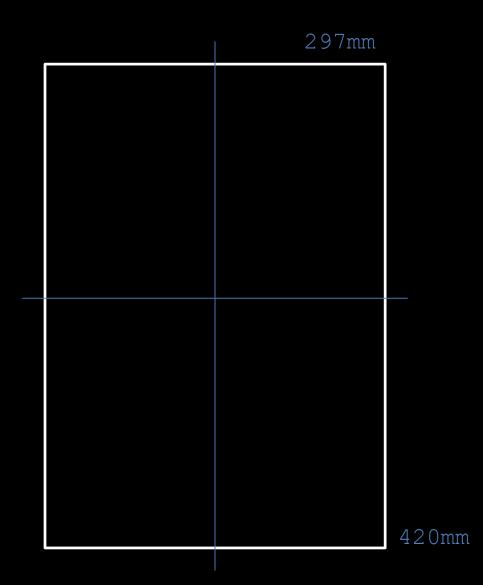


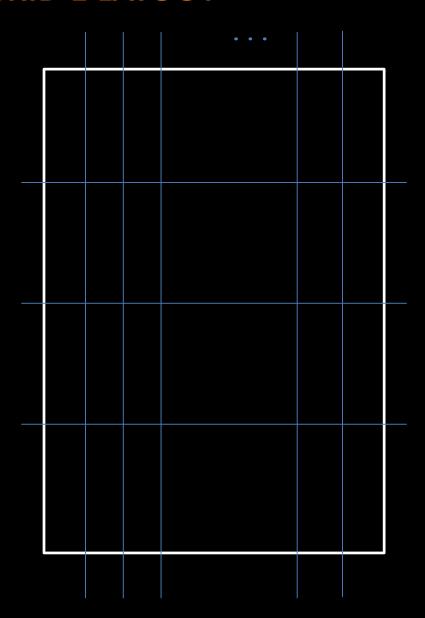
RStudioLogo.png





420mm





```
#ENTRADA DE DADOS
y1 \leftarrow round(rnorm(n = 36, mean = 7, sd = 2))
y2 < - round(rnorm(n = 36, mean = 21, sd = 6))
y3 < - round(rnorm(n = 36, mean = 50, sd = 8))
x \leftarrow sample(c(LETTERS[1:12], LETTERS[1:3]), 36, rep=T)
grp < -rep(c("G1", "G2", "G3"), each = 12)
dat \leftarrow data.frame(grp, x, y1, y2, y3)
tab <- data.frame(Freq=sort(table(x)))
#INICIANDO DIRETÓRIO DE TRABALHO
setwd("...")
#IMPORTANDO IMAGENS
require(png)
require (jpeq)
imag1 <- readPNG("bk.png")</pre>
im1<- rasterGrob(imag1, width = unit(297, "mm"), height = unit(420, "mm"))</pre>
imag2 <- readPNG("RStudioLogo.png")</pre>
im2 <- rasterGrob(imag2, width = unit(90, "mm"), height =unit(90, "mm"))
#CRIANDO IMAGEM
png("painel.png", width = 297 , height = 420, units = "
#CONSTRUIR UM NOVO GRID
grid.newpage()
```

```
#CABECALHO E TEXTOS
pushViewport(viewport(layout = grid.layout(1, 1)))
print(grid.draw(im1))
grid.text("MATE56 2021.1", x = unit(297/2+35, "mm"), y = unit(390, "mm"),
          gp = gpar(fontfamily = "VertigoFLF", col = cor.3, cex = 7))
grid.text("PROF. ANDERSON ARA", x = unit(297/2+35, "mm"), y = unit(360, "mm"),
          qp = qpar(fontfamily = "VertigoFLF", col = cor.2, cex = 5))
grid.text("DESt - UFBA", x = unit(297/2+20, "mm"), y = unit(335, "mm"), hjust=-0.1,
          qp = qpar(fontfamily = "VertigoFLF", col = cor.3, cex = 5))
grid.text("Salvador, 06 de maio de 2021", x = unit(297/2+65, "mm"), y = unit(10, 10)
"mm"), hjust=0.1,
          gp = gpar(fontfamily = "VertigoFLF", col = cor.3, cex = 3))
```

#INCLUINDO GRÁFICOS

```
pushViewport(viewport(layout = grid.layout(4, 14)))
print(g1, vp = viewport(layout.pos.row=2:3, layout.pos.col = 2:7))
print(g2, vp = viewport(layout.pos.row=2, layout.pos.col = 8:13))
print(g3, vp = viewport(layout.pos.row=3, layout.pos.col = 8:13))
print(g4, vp = viewport(layout.pos.row=4, layout.pos.col = 2:13))
```

#INCLUINDO RLOGO

```
pushViewport(viewport(layout.pos.row=1, layout.pos.col = 2:7))
print(grid.draw(im2))
```

dev.off()



MATE56 2021.1

PROF. ANDERSON ARA

DESt - UFBA

