**ATIVIDADE 2**

**EXERCÍCIOS**

1. Escolha 4 pessoas, colete as informações abaixo e crie um conjunto de dados contendo as seguintes variáveis:

* Sexo
* Idade
* Altura
* Estado Civil
* Há quantos anos terminou a graduação

Obs.: Escolha uma das formas de entrada de dados apresentadas na aula

* Construa um histograma para a variável Altura, um gráfico de colunas para Estado Civil e um gráfico de Dispersão para “Idade x Há quantos anos terminou a graduação”.
* Insira cores, legendas e títulos nos eixos dos gráficos.

HISTOGRAM

ggplot(data = dados) + geom\_histogram(aes(x=Altura, fill='Altura'), color='red', fill='yellow') + labs(fill='Altura') + theme(legend.position = "top")

COLUNAS

ggplot(dados, aes(x=as.factor(EstCivil), fill=as.factor(EstCivil))) + geom\_bar() + labs(fill='Estado Civil') + theme(legend.position = "top")

DISPERSÃO

ggplot(dados, aes(x=as.factor(Idade),y=as.factor(Grad))) + geom\_point()

1. Com base nos dados da tabela abaixo construa um gráfico de barras que mostra proporção de especialistas, mestres e doutores em relação ao total de docentes. Coloque cores, legenda, nomes nos eixos, etc.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Universidades** | **Especialistas** | **Mestres** | **Doutores** | **TOTAL** |
| Privadas | 210 | 355 | 482 | 1047 |
| Públicas Estaduais | 48 | 354 | 557 | 959 |
| Públicas Federais | 36 | 390 | 580 | 1006 |
| TOTAL | 294 | 1099 | 1619 | 3012 |

dados=read.csv("graduacao.csv", header = T, sep = ";")

categorias=c('especialistas','mestres', 'doutores')

#especialista = 294 / 3012 = 0.09760956 \* 100 = 9.7 %

#mestres = 1099 / 3012 = 0.3648738 \* 100 = 36.48 %

#doutores = 1619 / 3012 = 0.5375166 \* 100 = 53.75 %

#9.7 + 36.48 + 53.75 = 99.93 %

porcentagem=c(9.7,36.48,53.75)

resultado = matrix(c(categorias,porcentagem), nrow = 2, ncol = 3, byrow = T )

df = data.frame(

Categorias = categorias,

Porcentagem = porcentagem

)

ggplot(df, aes(x = Categorias, y = Porcentagem, fill = Categorias)) + geom\_bar(stat = 'identity') +

theme\_minimal()

1. Foram levantados os tempos (em horas por semana) que 20 alunos de uma escola da rede pública gastam na Internet e praticando atividades físicas. Os resultados deste levantamento são apresentados a seguir:

Aluno 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Internet 30 8 20 15 12 4 40 25 20 14 22 2 15 18 30 20 24 15 8 8

At. Físicas 2 10 5 5 8 15 0 4 5 10 2 15 6 4 2 6 4 10 12 15

Construa um diagrama de dispersão para esse par de variáveis. Com base neste gráfico, você acredita que há relação entre as horas dedicadas às duas atividades?

row\_1 = read.table(text = clipr::read\_clip())

row\_2 = read.table(text = clipr::read\_clip())

row\_3 = read.table(text = clipr::read\_clip())

dados\_df = data.frame(

Variavel1 = row\_1,

Variavel2 = row\_2

)

ggplot(dados\_df, aes(x = row\_1, y = row\_2)) +

geom\_point() +

geom\_smooth(method = "lm", se = FALSE) +

labs(x = "Variável 1", y = "Variável 2", title = "Diagrama de Dispersão")

Sim pois quanto mais tempo dedicado a internet menos atividade física é praticado