



# Formação Cientista de Dados

Configuração do Ambiente

# Configuração do Ambiente para R

---



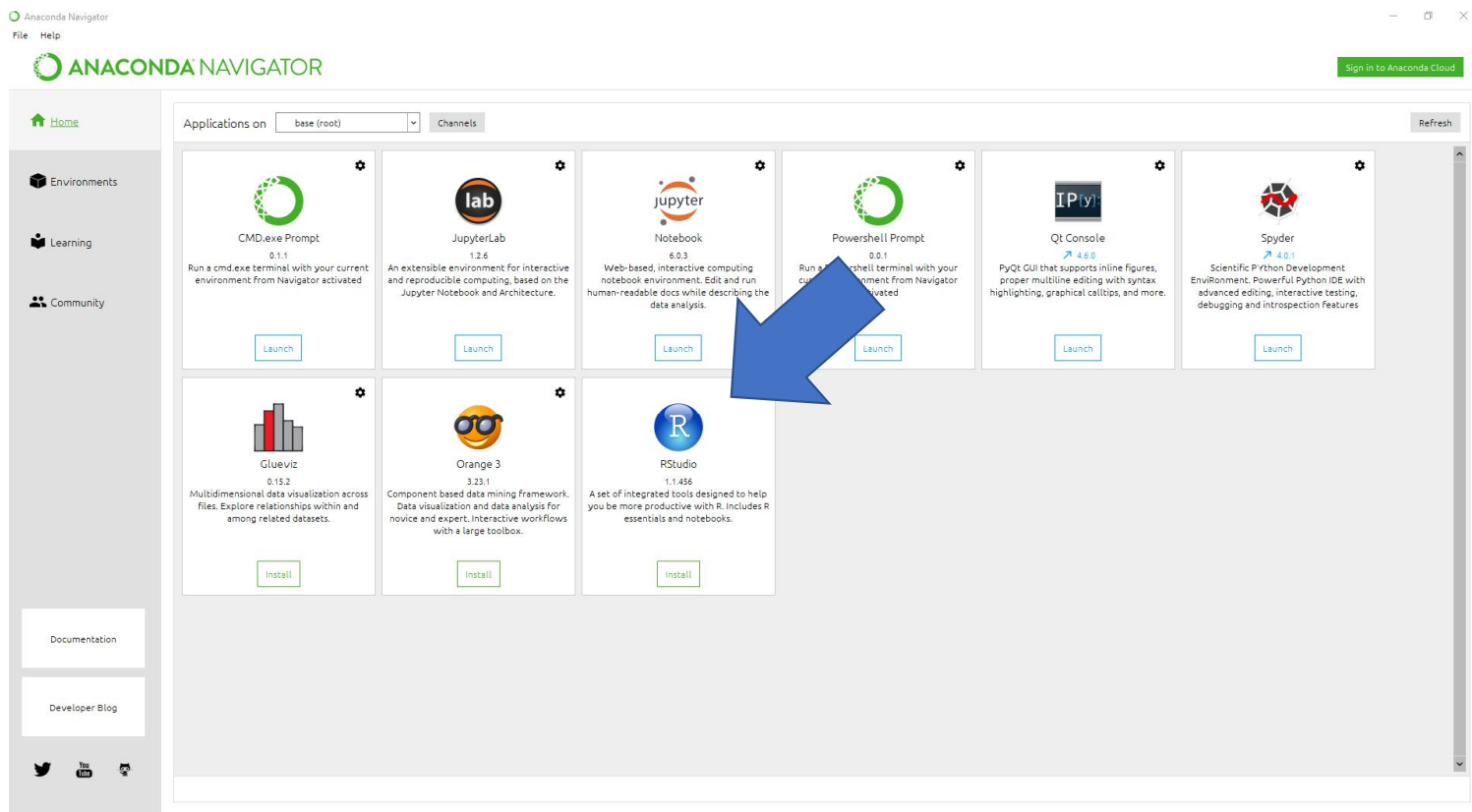
**Importante: Curso de Fundamentos  
do R**



**Você vai precisar baixar:**

R  
RStudio

# Opcional



# Códigos Fontes

- Notebook do R
- Extensão \*.Rmd
- Abre automaticamente no RStudio
- Arquivos de Dados devem estar no mesmo diretório do arquivo \*.Rmd

# Notebook do R

- Código é dividido em blocos
- Você pode rodar linha a linha (Ctrl + Enter)
- Você pode rodar todo o bloco (Ctrl + Shift + Enter)
- Em alguns casos você deve executar todo o bloco (Ctrl + Shift + Enter)



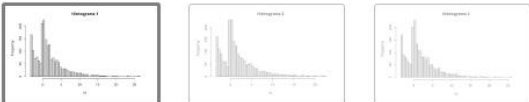


Untitled1\*6.Teorema central do limite.Rmd

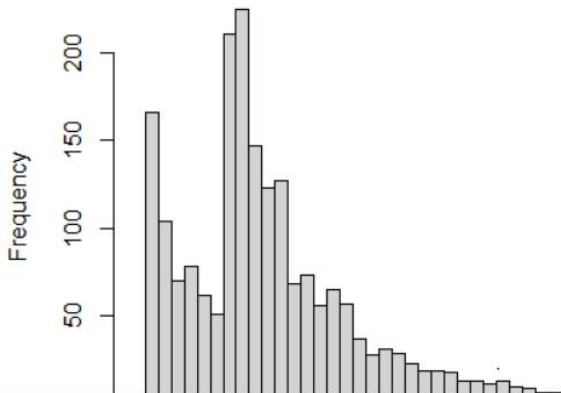
Insert

Run

```
20 for (i in 1:500) {
21   m = mvrnorm(1000, c(1, 2), matrix(c(10, 2, 2, 5), 2, 2),
22             skewness = c(5, 2), kurtosis = c(3, 3))
23   #grava a média da amostra
24   z[i] =mean(m)
25
26   #imprime as 3 primeiras
27   if (i<4){
28
29     hist(m,breaks=50, main= paste0("Histrograma ", i ))
30   }
31 }
32
```



### Histrograma 1



15:1

Chunk 2

R Markdown

ConsoleTerminalJobs

D:/Cursos Online/Em produção/Atualização CD/6.Estatística I/3. R/

# Instalação de Pacotes

Untitled1\* x 5.AprendizadoInstancia.Rmd x 4.SelecaoAtributos.Rmd x

Insert

↑

↓

Run

↺

⋮

```
1 ---
2 Formação Cientista de Dados - Prof. Fernando Amaral
3 Seleção de Atributos
4 ---
5
6 ```{r}
7 #install.packages("e1071")
8 #install.packages("randomForest")
9 library(e1071)
10 library(randomForest)
11 credito = read.csv("Credit.csv")
12 credito$class = as.factor(credito$class)
13
14 Dividimos os dados em treino e teste
15 Definimos semente para ser possível repetir o experimento
16 ```{r}
17 set.seed(234)
18 amostra = sample(2,1000,replace=T, prob=c(0.7,0.3))
19 creditotreino = credito[amostra==1,]
20 creditoteste = credito[amostra==2,]
21
22 Criamos um primeiro modelo com todos os atributos
23 Avaliamos a acurácia
24 ```{r}
25 modelo = svm(class ~., creditotreino)
26 predicacao = predict(modelo,creditoteste)
27 confusao = table(creditoteste$class,predicacao)
28 taxaacerto = (confusao[1] + confusao[4]) / sum(confusao)
29 taxaacerto
30
31 Aplicamos um método de seleção de atributos
32 ```{r}
33 importancia = randomForest(class ~ ., data = creditotreino)
34 col = importance(importancia)
35 col
36 varImpPlot(importancia)
37
38 Criamos um segundo modelo com as variáveis independentes mais importantes
39 ```{r}
40 modelo = svm(class ~ credit_amount + age + duration + checking_status|,creditotreino)
41 predicacao = predict(modelo,creditoteste)
42 confusao = table(creditoteste$class,predicacao)
43 taxaacerto = (confusao[1] + confusao[4]) / sum(confusao)
44 taxaacerto
45
46 ```
```

40:70 Chunk 5 R Markdown