

# MC536 – Bancos de Dados: Teoria e Prática

## Etapa 4

Gabriel Previato de Andrade - 172388

Ricardo Ribeiro Cordeiro - 186633

Grupo BolsoLula

# Entendendo os Dados

---

- *Heart Disease Data Set*  
(<https://archive.ics.uci.edu/ml/datasets/heart+Disease>)
- Hungarian Institute of Cardiology. Budapest
- University Hospital, Zurich, Switzerland
- University Hospital, Basel, Switzerland
- V.A. Medical Center, Long Beach and Cleveland Clinic Foundation

# Entendendo os Dados

---

- 76 atributos
- Muitos atributos não preenchidos
- Falta de padronização entre os centros médicos

# Entendendo os Dados

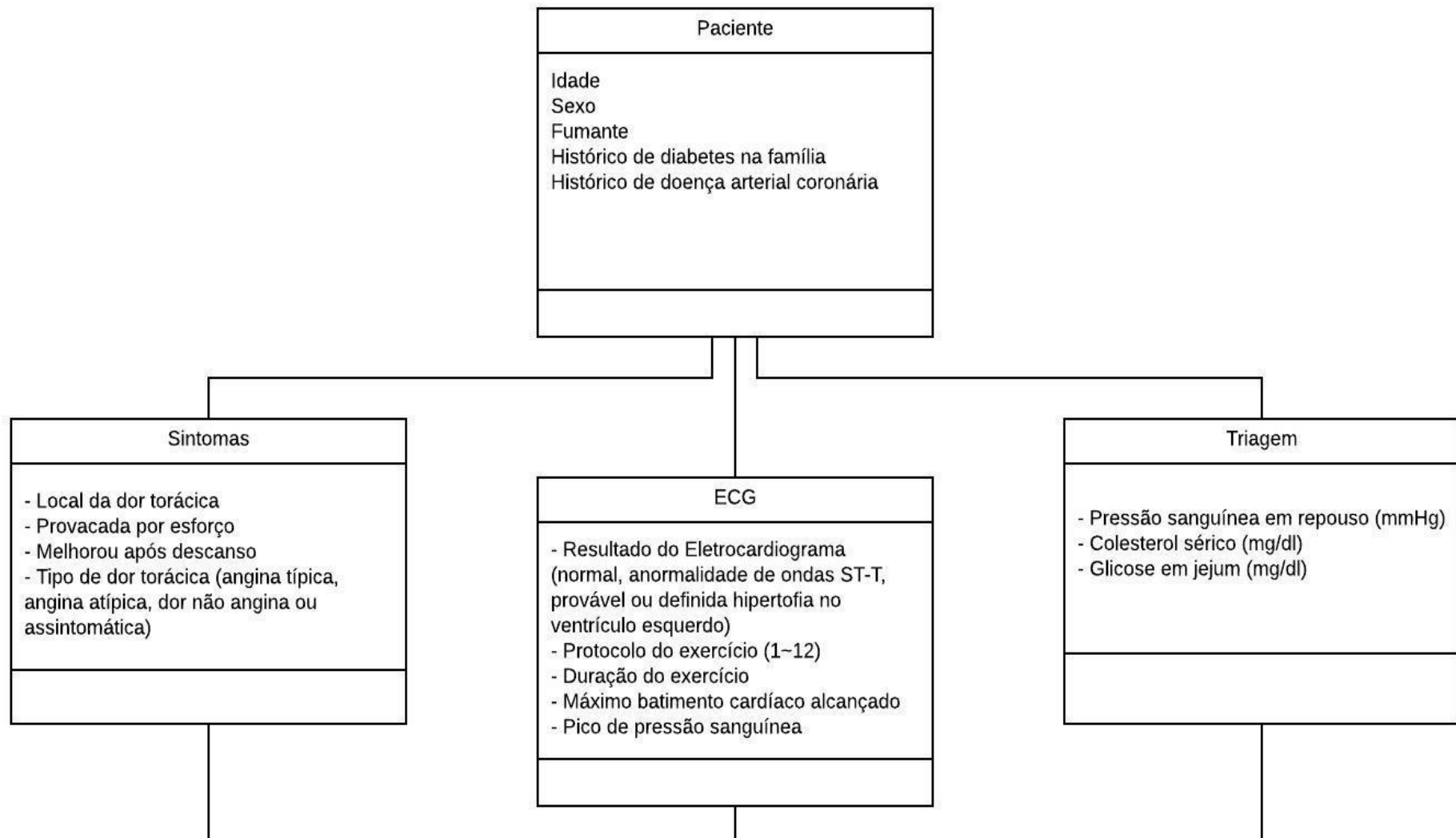
---

- Utilizamos 14 dos 76 atributos
- Apenas os dados do V.A. Medical Center, Cleveland

# Modelo Conceitual

---

- Como inicialmente pensamos:



Diagnóstico de doença cardíaca - saída

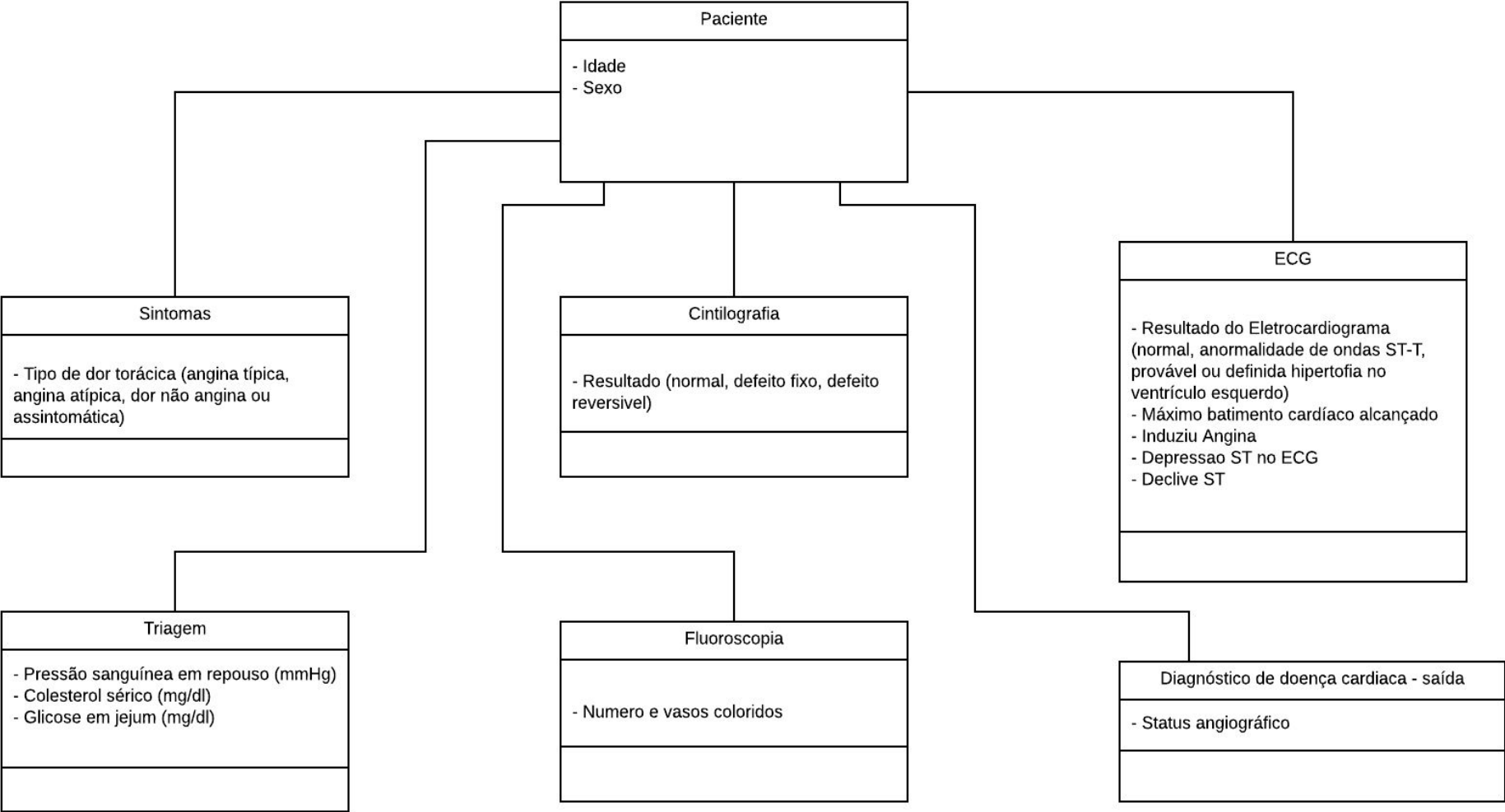
- Status angiográfico
- Possíveis vasos principais atingidos (1~9)

# Modelo Conceitual

---

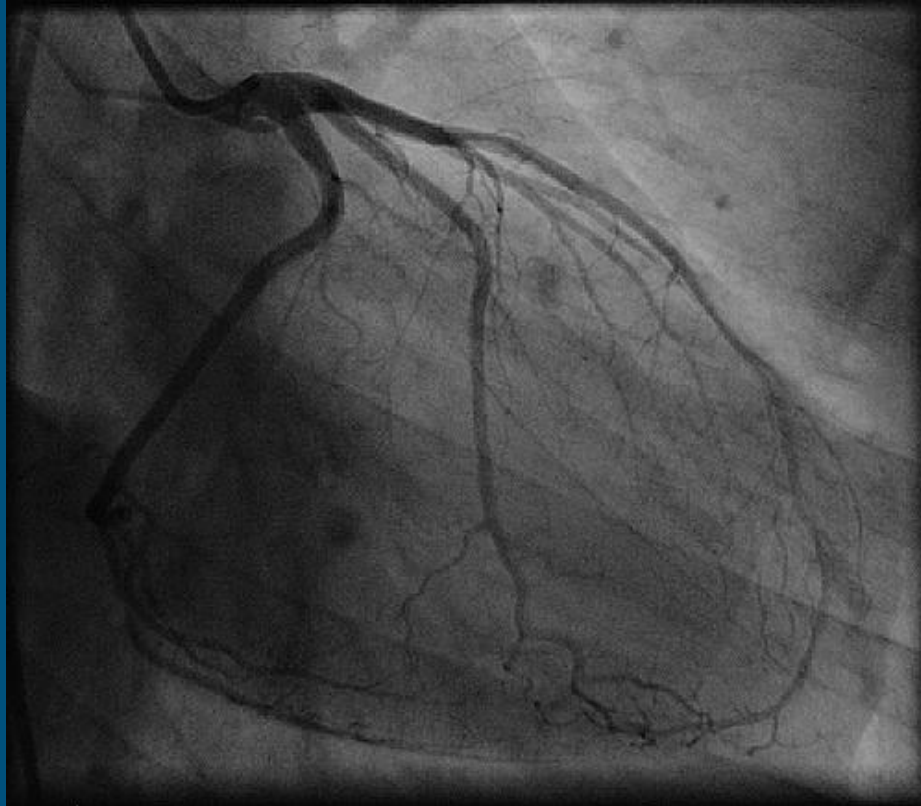
- Após mudanças dado os dados não preenchidos:





# Análise - Vasos coloridos pela Fluoroscopia

---



# Análise - Vasos coloridos pela Fluoroscopia

- 73% dos pacientes com nenhum vaso vaso colorido não apresentam doenças cardíacas.

```
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico = 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.numero_vasos = 0);  
  
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico > 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.numero_vasos = 0);
```

130

46

# Análise - Vasos coloridos pela Fluoroscopia

---

- 73% dos pacientes com nenhum vaso vaso colorido não apresentam doenças cardíacas.

```
LET $doc:= doc('mydoc.xml')
```

```
RETURN $doc//pessoa[diagnostico/diagnostico=0][fluoroscopia/numero_vasos=0]
```

```
< Pessoa>
  < idade> 63.0 </ idade>
  < genero> 1.0 </ genero>
  < uid> 6eaa2106-edce-11e8-a8c2-e03f49e51285 </ uid>
  < sintoma>
    < tipo_dor_toracica> 1.0 </ tipo_dor_toracica>
  </ sintoma>
  < triagem>
    < pressao_repouso> 145.0 </ pressao_repouso>
    < colesterol> 233.0 </ colesterol>
    < glicose> 1.0 </ glicose>
  </ triagem>
  < ecg>
    < resultado> 2.0 </ resultado>
    < batimento_max> 150.0 </ batimento_max>
    < induziu_angina> 0.0 </ induziu_angina>
    < depressao_st> 2.3 </ depressao_st>
    < declive_st> 3.0 </ declive_st>
  </ ecg>
  < fluoroscopia>
    < numero_vasos> 0.0 </ numero_vasos>
  </ fluoroscopia>
  < cintilografia>
    < thal> 6.0 </ thal>
  </ cintilografia>
  < diagnostico>
    < diagnostico> 0 </ diagnostico>
  </ diagnostico>
</ Pessoa>
```

# Análise - Vasos coloridos pela Fluoroscopia

- 82% dos pacientes com mais de 2 vasos coloridos apresentam doenças cardíacas.

```
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico = 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.numero_vasos >= 2);  
  
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico > 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.numero_vasos >= 2);
```

10

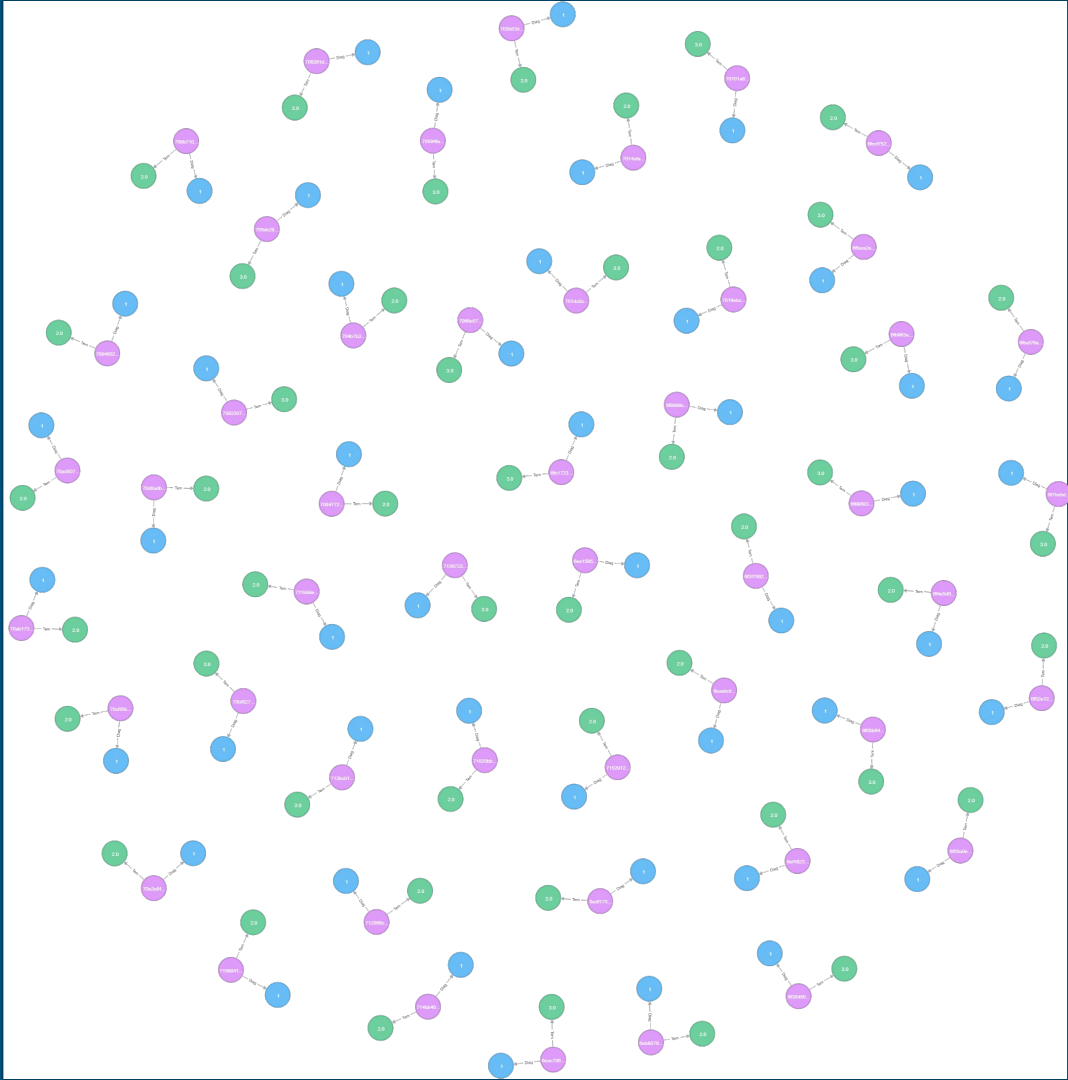
48

# Análise - Vasos coloridos pela Fluoroscopia

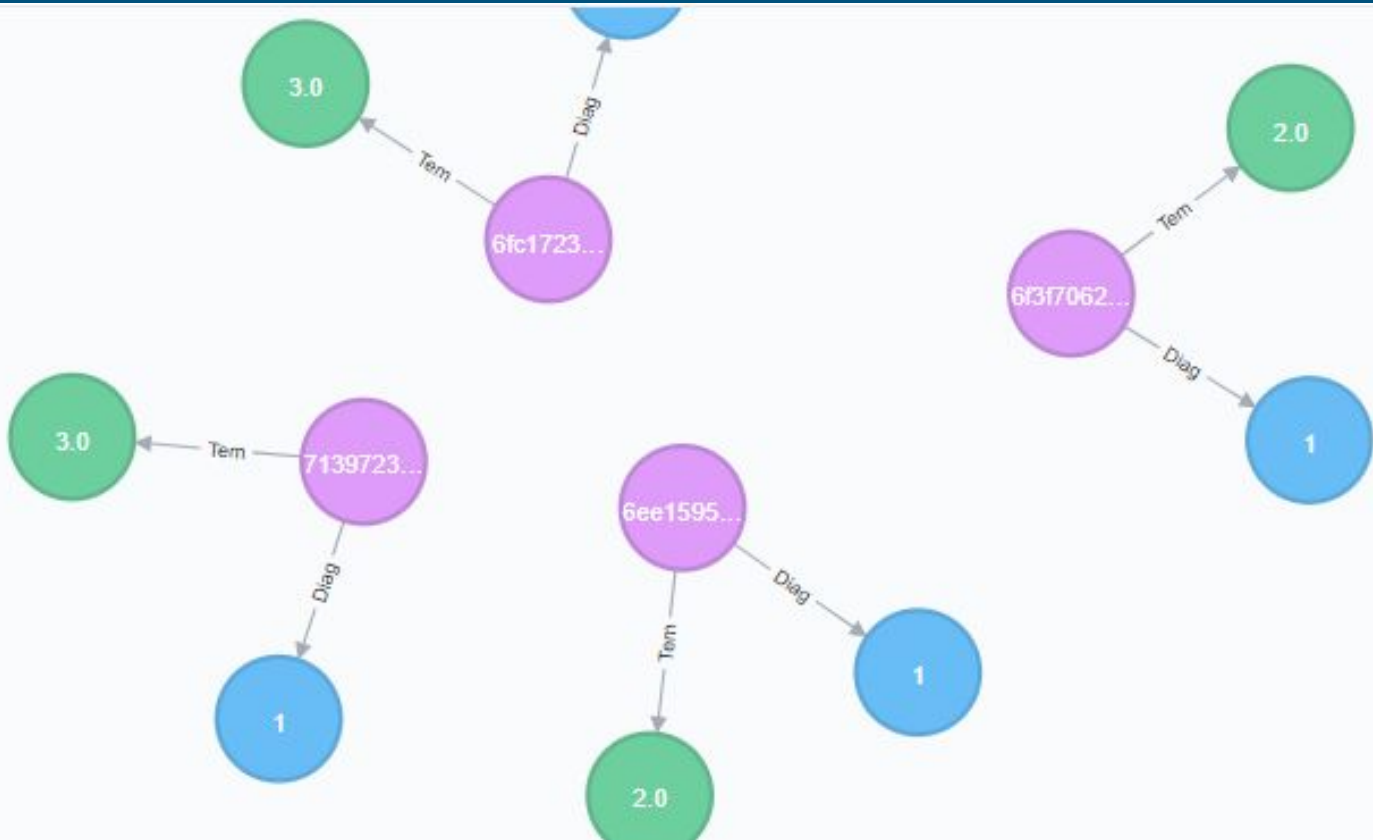
---

- 82% dos pacientes com mais de 2 vasos coloridos apresentam doenças cardíacas.

```
MATCH (p:Pessoa)-[]->(f:Fluoroscopia)
WHERE f.numero_vasos >= 2
MATCH (p:Pessoa)-[:Diag]->(d:Diagnostico {diagnostico : 0})
RETURN p, d, f
```



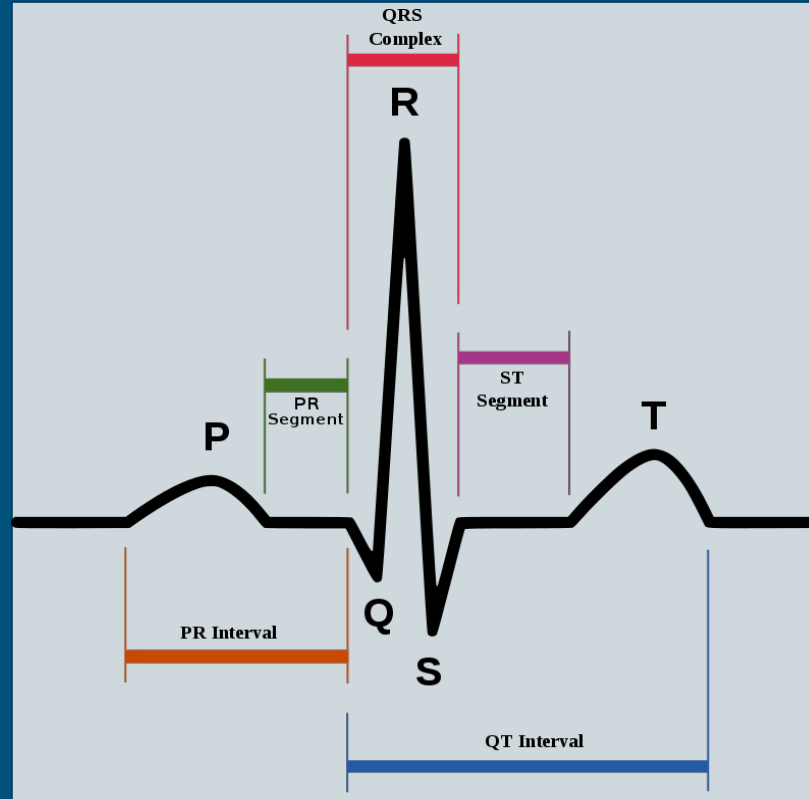




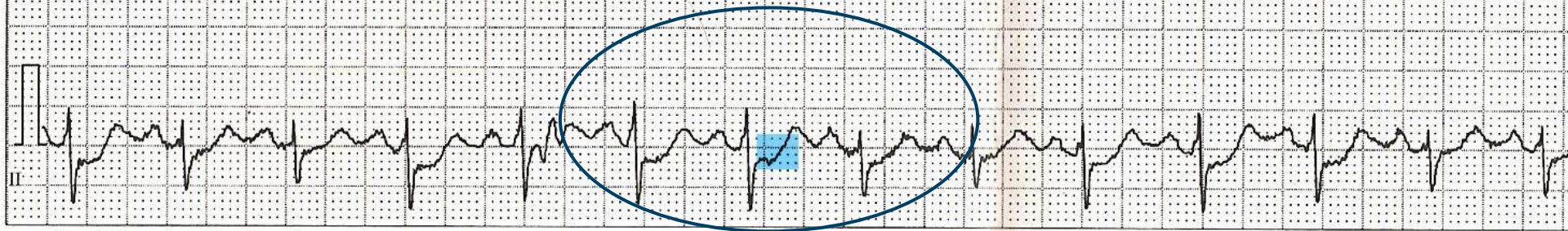
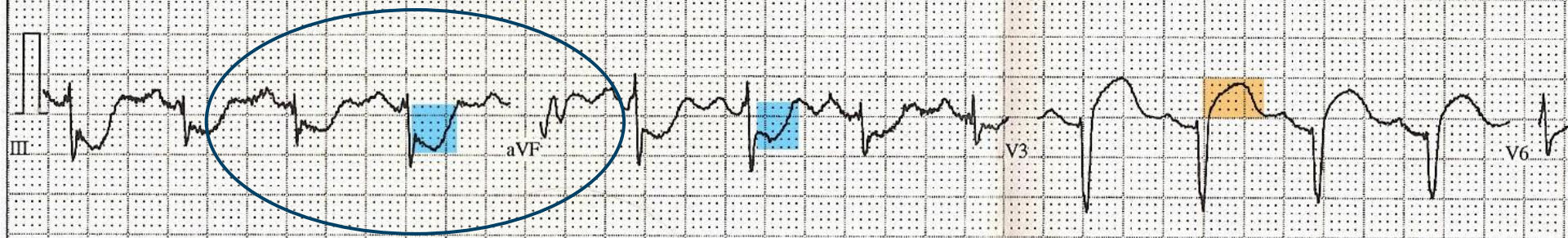
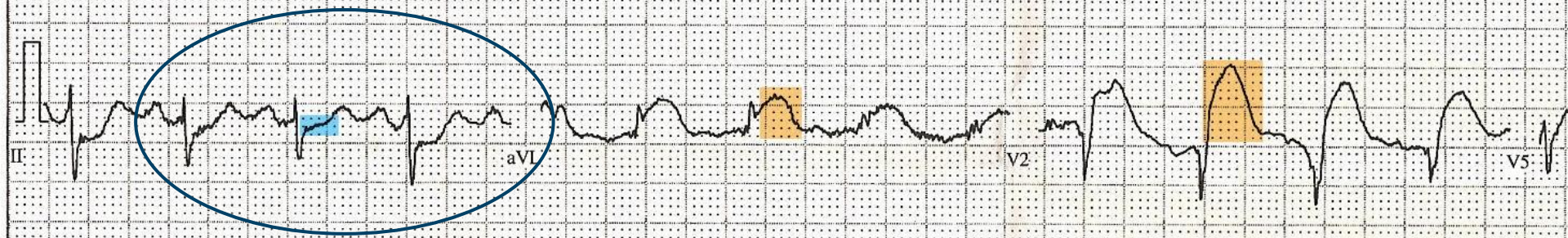
# Análise - Vasos coloridos pela Fluoroscopia



# Análise - Depressão do ST induzido por exercício







# Análise - Depressão do ST induzido por exercício

---

- 90% dos pacientes com depressão maior que 3 apresentam doença cardíaca.

```
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico = 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.depressao_st > 3);  
  
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico > 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.depressao_st > 3);
```

2

18



## Análise - Depressão do ST induzido por exercício

---

- 90% dos pacientes com depressão maior que 3 apresentam doença cardíaca.

```
let $doc:= doc('mydoc.xml')  
return $doc//pessoa[diagnostico/diagnostico=1][ecg/depressao_st>=3]
```

```
< Pessoa >
  < idade > 62.0 < / idade >
  < genero > 0.0 < / genero >
  < uid > 6eb80780-edce-11e8-aaf8-e03f49e51285 < / uid >
  < sintoma >
    < tipo_dor_toracica > 4.0 < / tipo_dor_toracica >
  < / sintoma >
  < triagem >
    < pressao_repouso > 140.0 < / pressao_repouso >
    < colesterol > 268.0 < / colesterol >
    < glicose > 0.0 < / glicose >
  < / triagem >
  < ecg >
    < resultado > 2.0 < / resultado >
    < batimento_max > 160.0 < / batimento_max >
    < induziu_angina > 0.0 < / induziu_angina >
    < depressao_st > 3.6 < / depressao_st >
    < declive_st > 3.0 < / declive_st >
  < / ecg >
  < fluoroscopia >
    < numero_vasos > 2.0 < / numero_vasos >
  < / fluoroscopia >
  < cintilografia >
    < thal > 3.0 < / thal >
  < / cintilografia >
  < diagnostico >
    < diagnostico > 1 < / diagnostico >
  < / diagnostico >
< / Pessoa >
```

# Análise - Depressão do ST induzido por exercício

- 74% dos pacientes com depressão menor que 0.6 não apresentam doença cardíaca.

```
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico = 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.depressao_st < 0.6);  
  
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, ECG E  
WHERE (D.diagnostico > 0 AND D.id_paciente = P.id) AND (E.id_paciente = P.id AND E.depressao_st < 0.6);
```

100

35



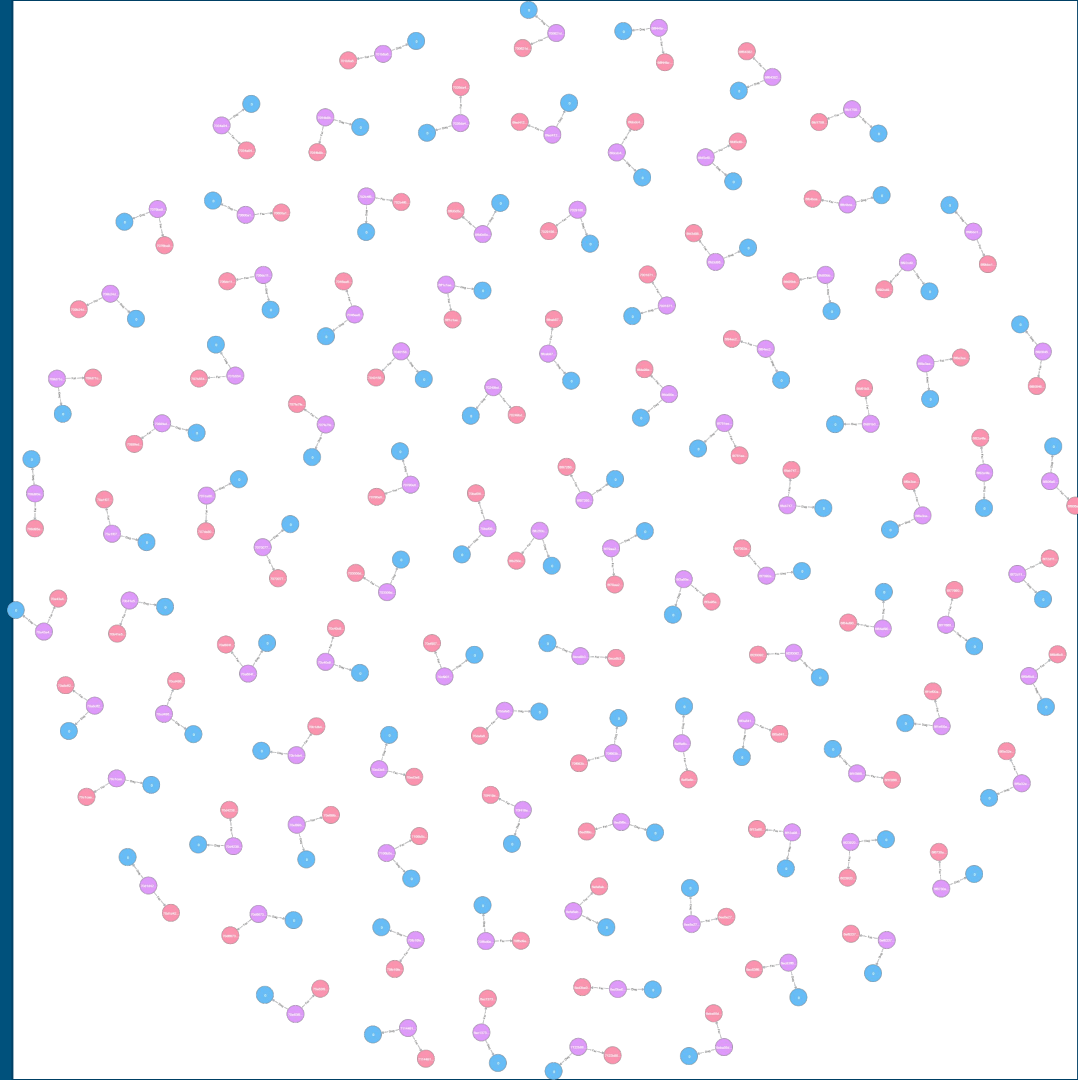
## Análise - Depressão do ST induzido por exercício

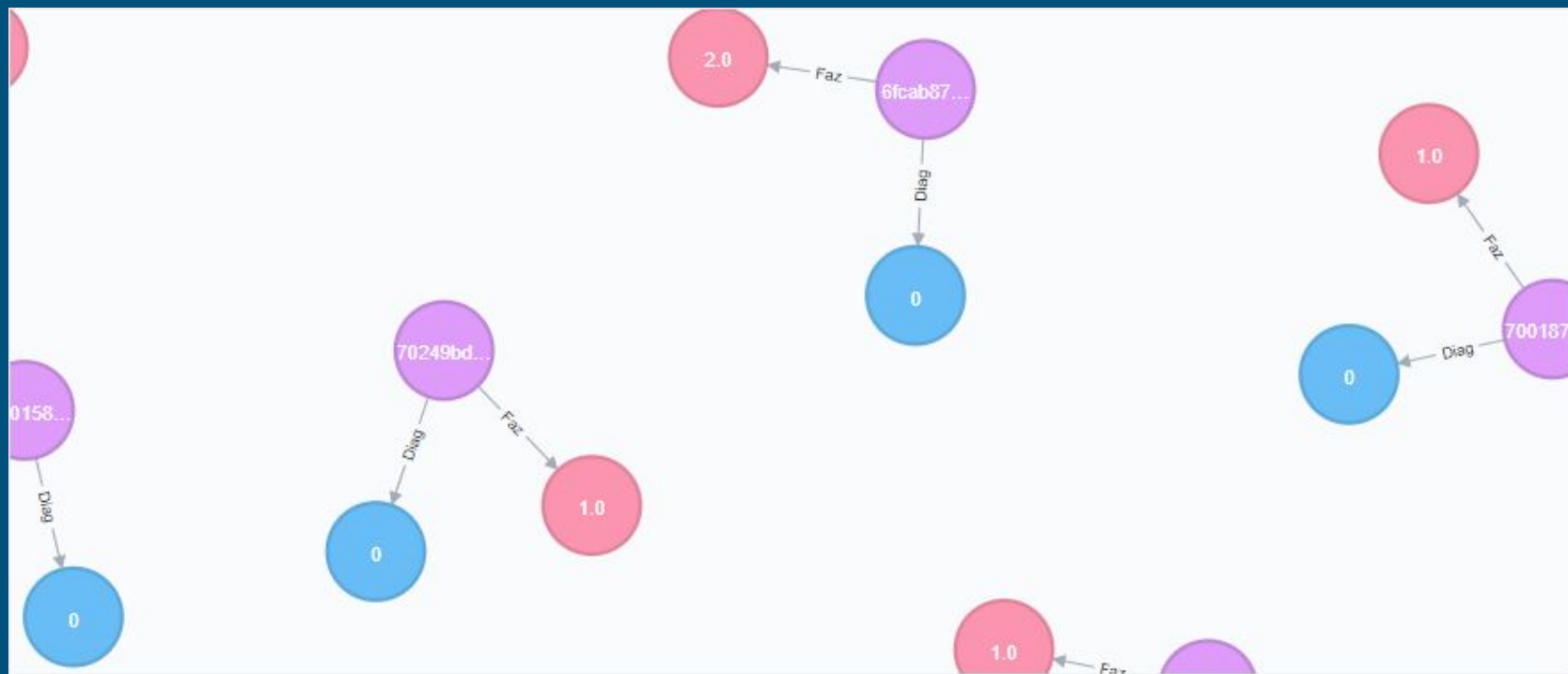
---

- 74% dos pacientes com depressão menor que 0.6 não apresentam doença cardíaca.

```
MATCH (p:Pessoa)-[]->(e:ECG)
WHERE e.depressao_st <= 0.6
MATCH (p:Pessoa)-[:Diag]->(d:Diagnostico {diagnostico : 0})
RETURN p, d, e
```

—





## Current relation

Relation: cleveland\_id\_norm-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,...  
Instances: 303

Attributes: 14  
Sum of weights: 303

## Attributes

All

None

Invert

Pattern

No.	Name
1	<input type="checkbox"/> idade
2	<input type="checkbox"/> genero
3	<input type="checkbox"/> tipo_dor_toracica
4	<input type="checkbox"/> pressao_repouso
5	<input type="checkbox"/> colesterol
6	<input type="checkbox"/> glicose
7	<input type="checkbox"/> resultado
8	<input type="checkbox"/> batimento_max
9	<input type="checkbox"/> induziu_angina
10	<input checked="" type="checkbox"/> depressao_st
11	<input type="checkbox"/> declive_st
12	<input type="checkbox"/> numero_vasos
13	<input type="checkbox"/> thal
14	<input type="checkbox"/> diagnostico

Remove

## Selected attribute

Name: depressao\_st  
Missing: 0 (0%)

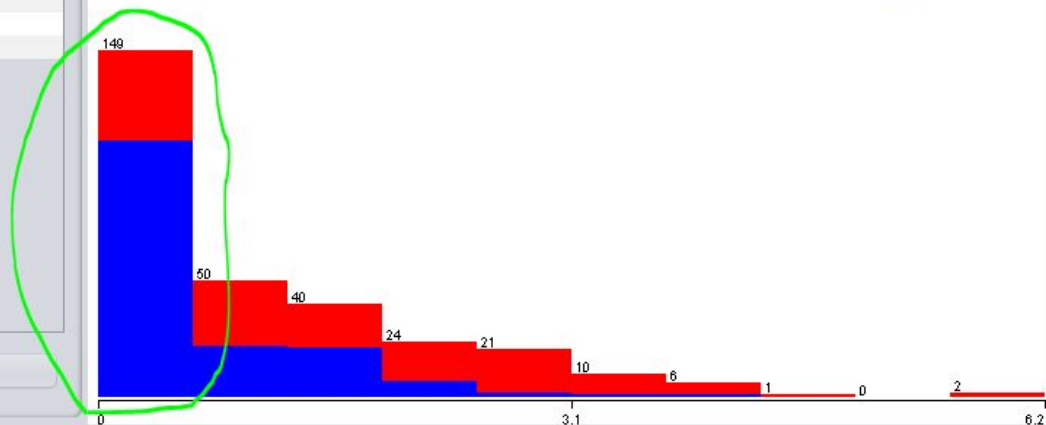
Distinct: 40

Type: Numeric  
Unique: 10 (3%)

Statistic	Value
Minimum	0
Maximum	6.2
Mean	1.04
StdDev	1.161

Class: diagnostico (Nom)

Visualize All



## Análise - Nível de glicose em jejum

---

- 54% dos pacientes com o nível de glicose menor que 120 mg/dl não apresentam doença cardíaca.
- 49% dos pacientes com o nível de glicose maior que 120 mg/dl apresentam doença cardíaca.

# Análise - Nível de glicose em jejum

---

- 54% dos pacientes com o nível de glicose menor que 120 mg/dl não apresentam doença cardíaca.

```
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, Triagem T  
WHERE (D.diagnostico = 0 AND D.id_paciente = P.id) AND (T.id_paciente = P.id AND T.glicose = 0);  
  
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, Triagem T  
WHERE (D.diagnostico > 0 AND D.id_paciente = P.id) AND (T.id_paciente = P.id AND T.glicose = 0);
```

141

117

# Análise - Nível de glicose em jejum

---

- 49% dos pacientes com o nível de glicose maior que 120 mg/dl apresentam doença cardíaca.

```
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, Triagem T  
WHERE (D.diagnostico = 0 AND D.id_paciente = P.id) AND (T.id_paciente = P.id AND T.glicose = 1);  
  
SELECT COUNT(*)  
FROM Diagnostico D, Paciente P, Triagem T  
WHERE (D.diagnostico > 0 AND D.id_paciente = P.id) AND (T.id_paciente = P.id AND T.glicose = 1);
```

23

22

# Análise utilizando Weka

---

- Analisar os dados visualmente
- Tentar criar modelos de predição de doença



# Current relation

Relation: cleveland\_id\_norm-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,...  
Instances: 303

Attributes: 14  
Sum of weights: 303

## Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> idade
2	<input type="checkbox"/> genero
3	<input type="checkbox"/> tipo_dor_toracica
4	<input type="checkbox"/> pressao_repouso
5	<input type="checkbox"/> colesterol
6	<input type="checkbox"/> glicose
7	<input type="checkbox"/> resultado
8	<input checked="" type="checkbox"/> batimento_max
9	<input type="checkbox"/> induziu_angina
10	<input type="checkbox"/> depressao_st
11	<input type="checkbox"/> declive_st
12	<input type="checkbox"/> numero_vasos
13	<input type="checkbox"/> thal
14	<input type="checkbox"/> diagnostico

Remove

## Status

## Selected attribute

Name: batimento\_max  
Missing: 0 (0%)

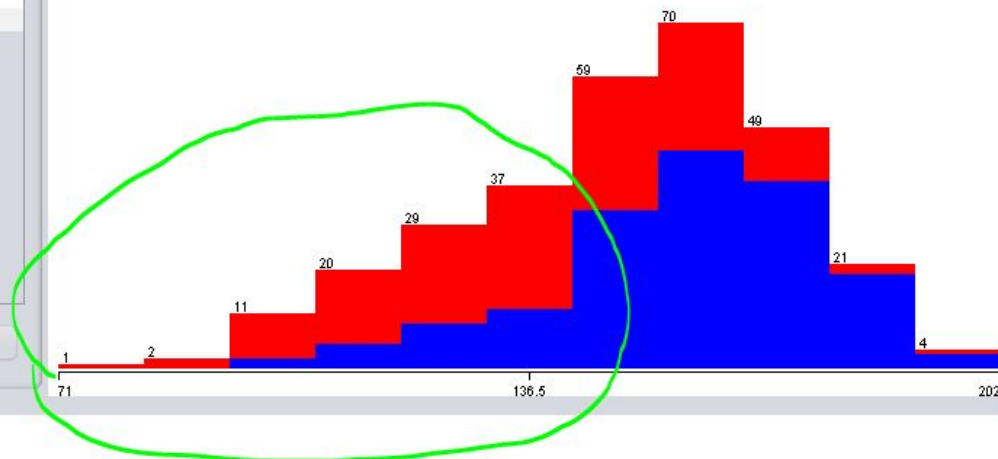
Distinct: 91

Type: Numeric  
Unique: 28 (9%)

Statistic	Value
Minimum	71
Maximum	202
Mean	149.607
StdDev	22.875

Class: diagnostico (Nom)

Visualize All

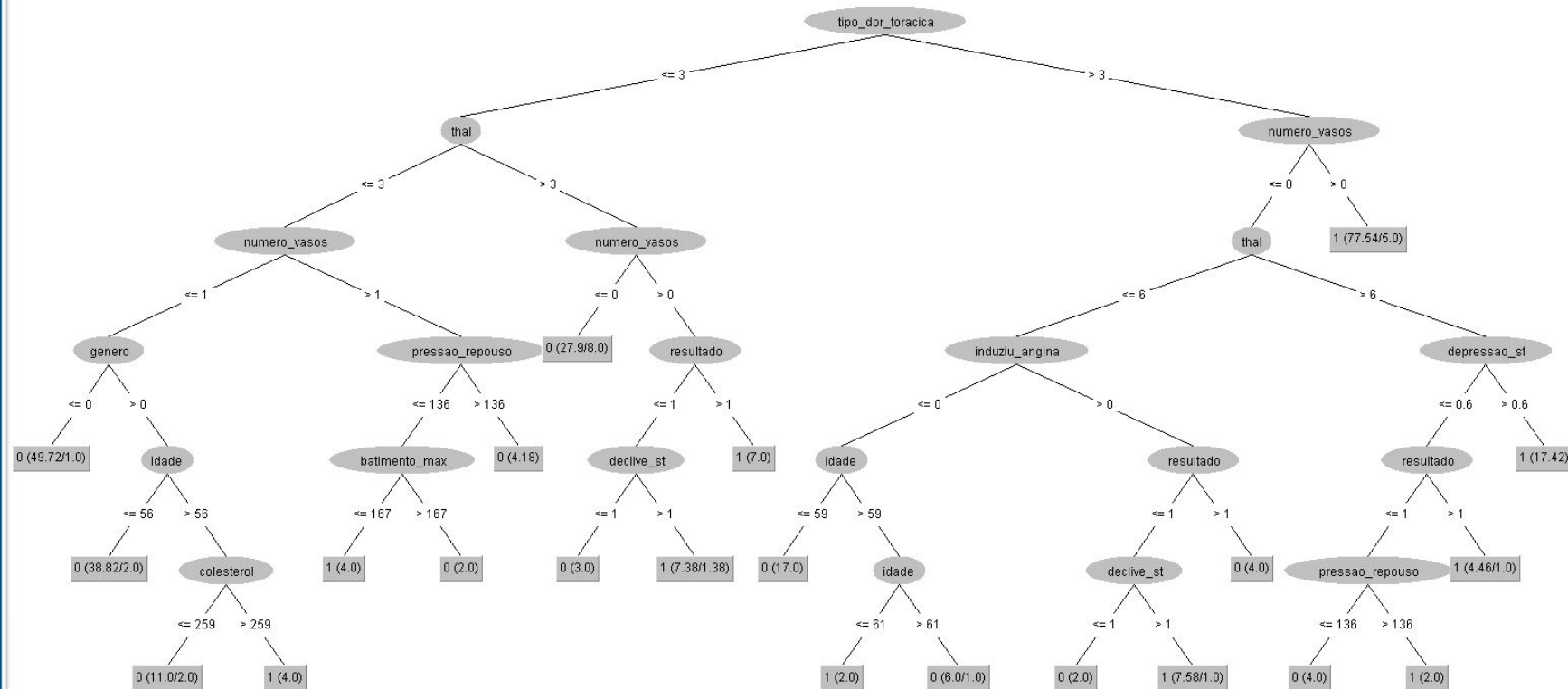


# Análise utilizando Weka

- Árvore J48

Correctly Classified Instances	238	78.5479 %
Incorrectly Classified Instances	65	21.4521 %
Kappa statistic	0.5664	
Mean absolute error	0.256	
Root mean squared error	0.4263	
Relative absolute error	51.5439 %	
Root relative squared error	85.5453 %	
Total Number of Instances	303	

Tree View



# Análise utilizando Weka

- Regressão Logística

Correctly Classified Instances	256	84.4884 %
Incorrectly Classified Instances	47	15.5116 %
Kappa statistic	0.6861	
Mean absolute error	0.226	
Root mean squared error	0.3466	
Relative absolute error	45.5134 %	
Root relative squared error	69.552 %	
Total Number of Instances	303	

# Análise utilizando Weka

Logistic Regression with ridge parameter of 1.0E-8  
Coefficients...

Variable	Class
idade	0.0133
genero	-1.3408
tipo_dor_toracica	-0.6127
pressao_repouso	-0.0236
colesterol	-0.0051
glicose	0.8064
resultado	-0.2464
batimento_max	0.0197
induziu_angina	-1.0212
depressao_st	-0.2634
declive_st	-0.5694
numero_vasos	-1.2276
thal	-0.3337
Intercept	7.6801