



MATEUS DO NASCIMENTO MAGALHÃES DA SILVA
FABRICIO ALMEIDA DA SILVA NUNES
ISMAEL WESLEY NEVES DE BRITO

**ANÁLISE DE DADOS DA VOTAÇÃO SOBRE A OPÇÃO DE MODALIDADE DE
AULA DOS ALUNOS DA UNICARIOCA DE 2022 ATRAVÉS DA LINGUAGEM R**

RIO DE JANEIRO

2022

MATEUS DO NASCIMENTO MAGALHÃES DA SILVA
FABRICIO ALMEIDA DA SILVA NUNES
ISMAEL WESLEY NEVES DE BRITO

**ANÁLISE DE DADOS DA VOTAÇÃO SOBRE A OPÇÃO DE MODALIDADE DE
AULA DOS ALUNOS DA UNICARIOCA DE 2022 ATRAVÉS DA LINGUAGEM R**

Atividade de conclusão da matéria de análise computacional e qualitativa de dados, apresentado ao Centro Universitário Carioca (Unicarioca), como requisito exigido parcial à obtenção do grau da matéria.

Docente(a): Prof. Manuel Martins Filho, PhD.

RIO DE JANEIRO

2022

RESUMO

Devido ao atual cenário vivenciado pela humanidade a partir de 2020, a pandemia do Covid-19, as universidades tornaram as aulas remotas, para poderem manter a continuidade das aulas. No ano de 2022 as aulas voltaram a ser realizadas presencialmente e remotamente, devido a isso, foi realizada uma pesquisa com os discentes do Centro Universitário Carioca, sobre qual modalidade eles prefeririam para estudar, remoto, híbrido ou presencial. Neste relatório foi utilizado a linguagem R como ferramenta de análise de dados, desde o carregamento quanto às consultas.

ABSTRACT

Because of the current scenario experienced by humanity from 2020, the Covid-19 pandemic, the universities made the classes remote, to be able to maintain continuity of classes. In the year 2022 the classes will be held again in person and remotely, so a survey was conducted with the students of the Centro Universitário Carioca, about which modality they would prefer to study, the options were remote, hybrid or local. In this report the R language was used as a tool for data analysis, from loading to queries.

INTRODUÇÃO

Houve um grande impacto com a pandemia do COVID 19 na vida estudantil dos discentes universitários desde de 2020. Na UniCarioca as aulas passaram a ser remotas durante a alta de contágio e morte da pandemia, todavia quando chegou o ano de 2022, as aulas retornaram para o modelo presencial.

Muitos alunos já se adaptaram ao modelo remoto, não somente a despeito da universidade, mas também para as suas atividades laborais.

Esse relatório possui a opinião pessoal dos alunos a respeito de suas preferências.

METODOLOGIA

O professor Manuel Filho realizou uma pesquisa com os seus alunos sobre as suas preferências de modalidade de aula. Os dados obtidos na pesquisa, foram realizados no final do 1º semestre de 2022 em 10 turmas (todas com aulas remotas) dos cursos de Informática e Engenharia da Unicarioca. O objetivo da pesquisa é mensurar a preferência desse grupo de alunos em relação ao modelo de aula para 2022.2.

EXPECTATIVAS:

Diretores da Instituição: informação sobre as preferências dos alunos em relação ao modelo de aula resultante da nova realidade. O modelo remoto foi utilizado durante 5 semestres pela instituição (2 anos e meio), assim. O resultado da pesquisa será considerado no momento da elaboração do calendário de turmas para 2022.2.

Cientistas de Dados: gerar um elenco de informações que agregue valor aos dados e permita a instituição tomar decisões mais assertivas.

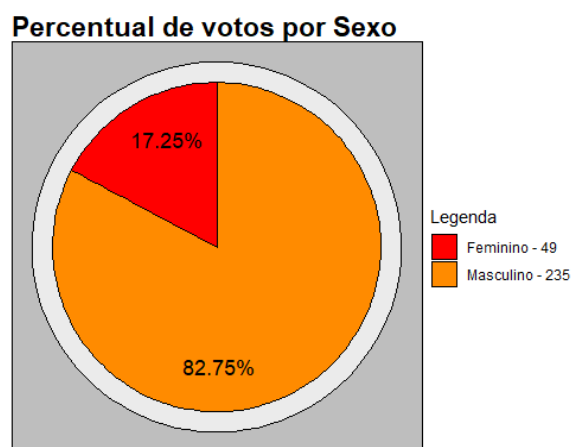
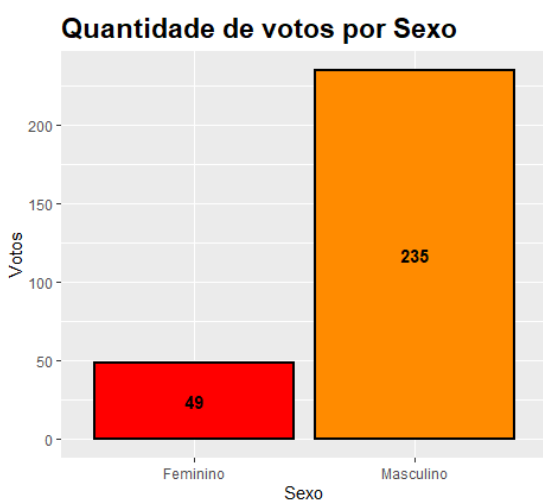
OBSERVAÇÕES IMPORTANTES:

- Os dados obtidos da pesquisa são de **457** alunos distintos, porém contém **173** respostas vazias que não serão úteis, logo para esta análise serão utilizadas **284** respostas.
- A amostra é de alunos dos cursos de **Informática e Engenharia**.
- Todas as turmas são de um **mesmo docente**.
- Todos os indicadores serão explorados, estão disponíveis as informações sobre: **disciplina, sexo, dia da semana, turno, tempo de aula** (1º tempo, 2º tempo, etc), **período da disciplina...**

ANÁLISES RELACIONADAS AO SEXO

A maioria dos alunos que votaram, são do sexo Masculino, com uma diferença percentual relativamente muito grande, de **82.75%** para apenas **17.25%**, sendo demonstrado na prática como a maioria dos alunos desses cursos são normalmente predominantes do sexo Masculino.

Sexo	Masculino	Feminino
QtdAlunos	235	49

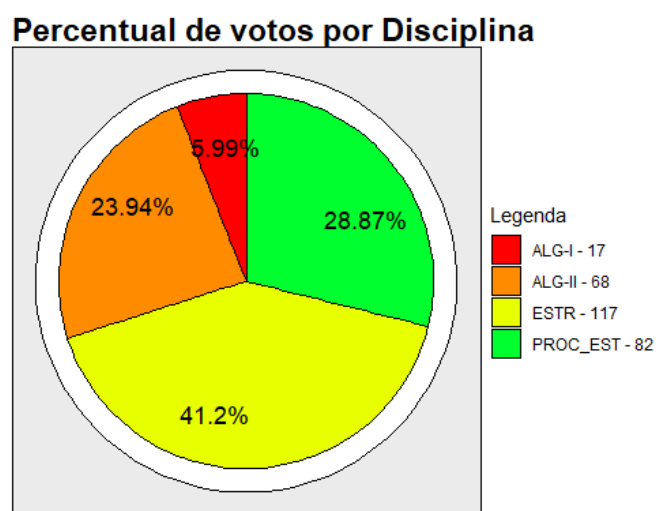
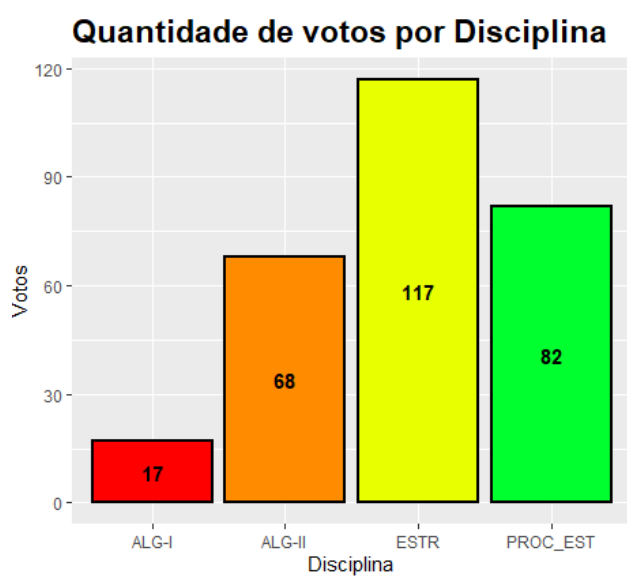


Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo	Variância(n)	Desvio(n)
49.0	95.5	142.0	142.0	188.5	235.0	8649	93

ANÁLISES RELACIONADAS A DISCIPLINA

A maior parte dos votos foram coletados da turma de Estrutura de dados (Terceiro período), tendo uma porcentagem total de **41.2%** dos votos, ou seja, é a disciplina que tem maior influência sobre o resultado final da coleta de votos.

Matéria	ALG-I	ALG-II	ESTR	PROC_EST
QtdAlunos	17	68	117	82

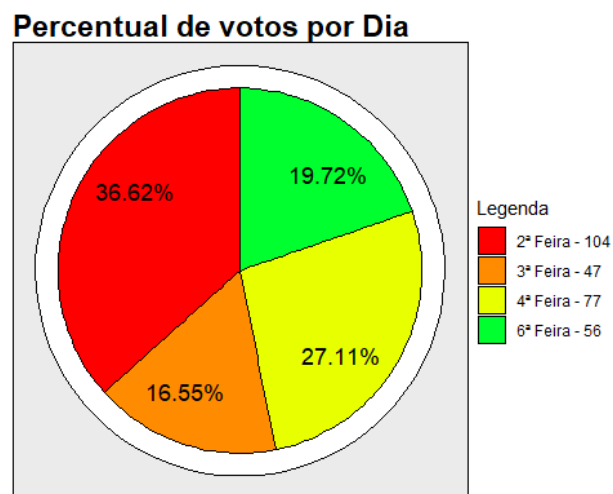
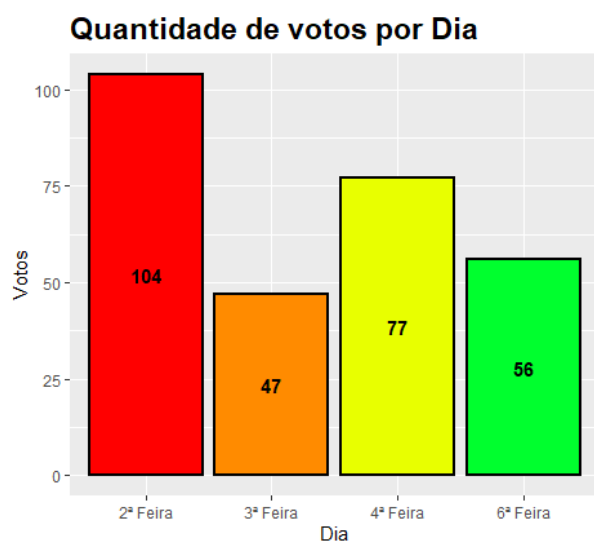


Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo	Variância(n)	Desvio(n)
17.00	55.25	75.00	71.00	90.75	117.00	1290.5	35.92353

ANÁLISES RELACIONADAS AO DIA DA SEMANA

A maior parte dos votos foram obtidos de alunos que estudavam em uma segunda-feira, sendo um total de 104 votos, ou seja, um percentual de **36.62%** dos votos totais coletados.

DiaSemana	2ª Feira	3ª Feira	4ª Feira	6ª Feira
QtdAlunos	104	47	77	56

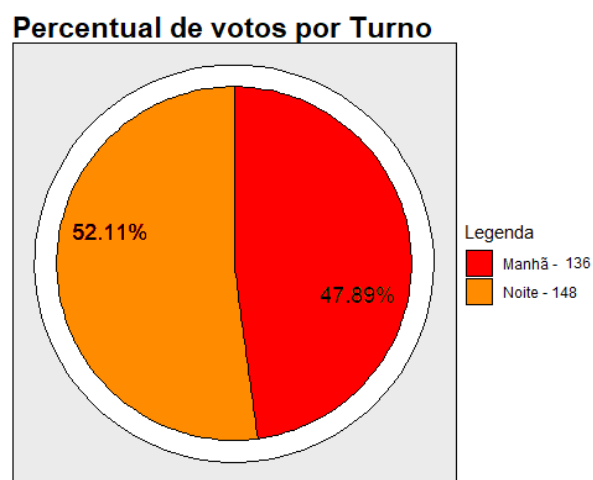
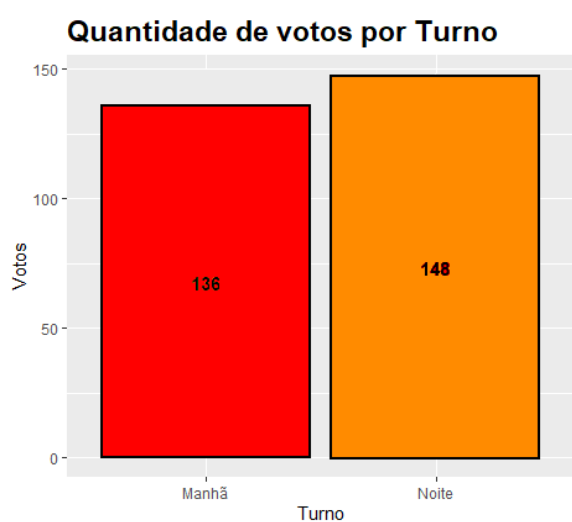


Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância(n)	Desvio(n)
47.00	53.75	66.50	71.00	83.75	104.00	481.5	21.94311

ANÁLISES RELACIONADAS AO TURNO

A diferença de turnos não foi tão evidente assim, mesmo a sua maioria tendo sido recolhida de alunos das turmas da noite (148 votos), sua diferença percentual está bem equilibrada, possuindo um percentual de **52.11%**, ou seja, pouco mais que a metade dos votos, influenciando um pouco mais do que os alunos da manhã.

Turno	Manhã	Noite
QtdAlunos	136	148



Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância(n)	Desvio(n)
136	139	142	142	145	148	36	6

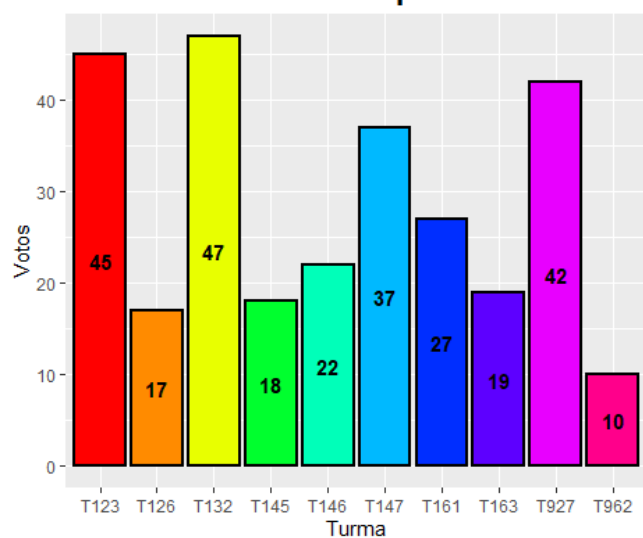
ANÁLISES RELACIONADAS A TURMA

Em relação às turmas, tem-se obtido 4 turmas que se destacaram com os seus votos, respectivamente as turmas: 132 (**16.55%**) 123 (**15.85%**) 927 (**14.79%**) e 147 (**13.03%**), as únicas turmas com mais de 10% dos votos cada, sendo as que possuem maior peso sobre os votos,.

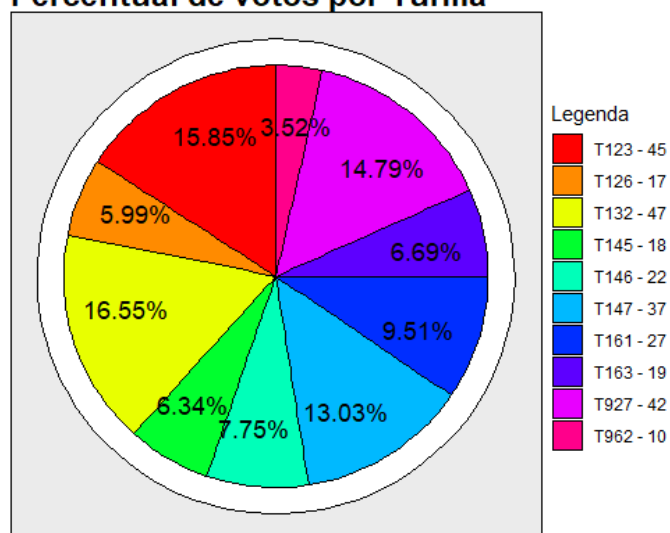
Já a turma com menos participação e relevância durante os resultados é a turma 962 com apenas 10 votos (**3.52%**), não influenciando muito nos votos.

Turma	123	126	132	145	146	147	161	163	927	962
QtdAlunos	45	17	47	18	22	37	27	19	42	10

Quantidade de votos por Turma



Percentual de votos por Turma

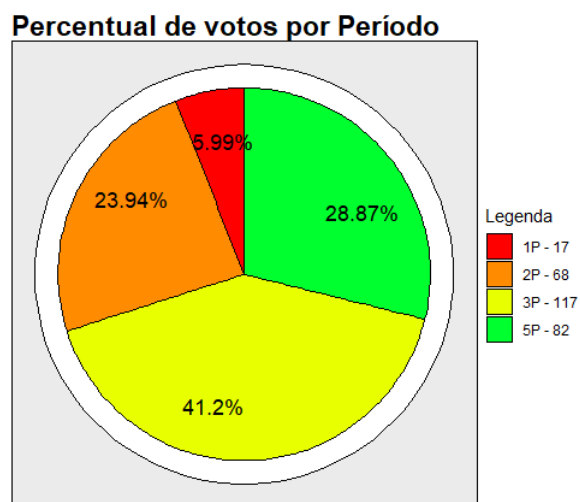
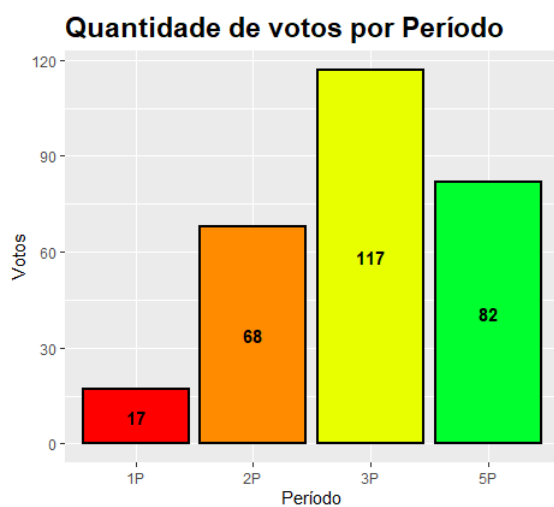


Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância(n)	Desvio(n)
10.00	18.25	24.50	28.40	40.75	47.00	158.84	12.60317

ANÁLISES RELACIONADAS AO PERÍODO

Em relação ao período dos alunos durante a pesquisa, a maior concentração de alunos, se encontra a partir do terceiro período, sendo esse o de maior quantidade de votos, tendo um total de 117 votos (41.2%) .

Período	1P	2P	3P	5P
QtdAlunos	17	68	117	82

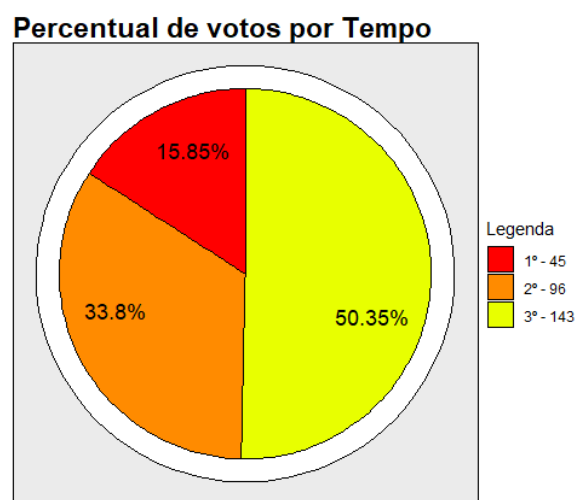
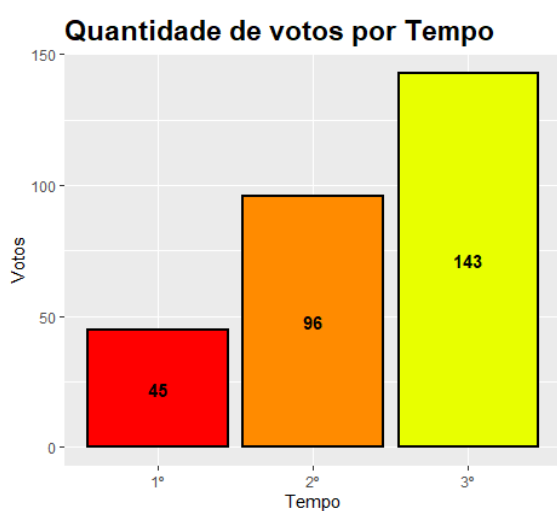


Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância(n)	Desvio(n)
17.00	55.25	75.00	71.00	90.75	117.00	1290.5	35.92353

ANÁLISES RELACIONADAS AO TEMPO DE AULA

Já, em relação ao tempo de aula dos alunos, quanto mais tarde o tempo de aula for, maior foi a sua concentração de votos, sendo os alunos do 3º tempo os que mais tiveram influência no resultado da votação com 143 votos (**50.35%**) e os que menos tiveram influência foram os alunos do 1º tempo com 45 votos (**15.85%**).

Tempo	1º	2º	3º
QtdAlunos	45	96	143

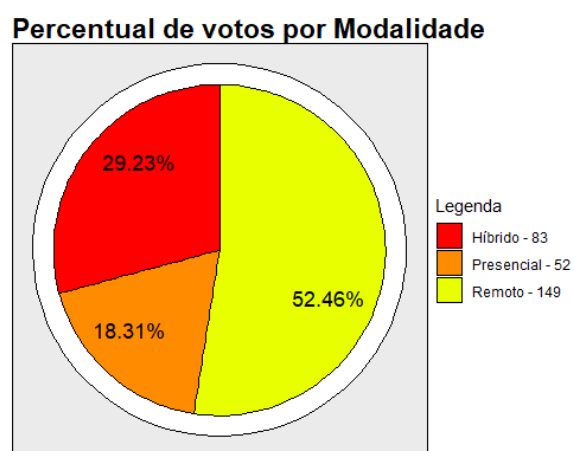
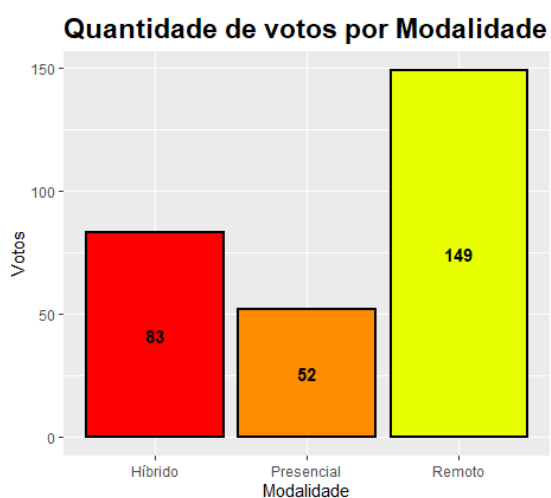


Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância(n)	Desvio(n)
45.00	70.50	96.00	94.67	119.50	143.00	1601.556	40.01944

ANÁLISES RELACIONADAS À MODALIDADE

Agora em relação à modalidade de ensino, obtém-se uma vitória a opção por ensino remoto, tendo um total de 149 votos (**52.46%**), ou seja mais do que a soma de seus concorrentes, híbrido com 83 e presencial com 52.

Modalidade	Híbrido	Presencial	Remoto
Votos	83	52	149



Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Variância(n)	Desvio(n)
52.00	67.50	83.00	94.67	116.00	149.00	1636.222	40.45024

PERGUNTAS PARA PESQUISAS EM RELAÇÃO AOS DADOS

As seguintes perguntas foram feitas para demonstrar como diversos tipos informações importantes sobre esses dados podem ser obtidos de forma fácil e intuitiva através da linguagem R e de suas ferramentas e funções:

1. Escolha dos alunos que estudam no primeiro tempo do turno da noite.

```
#table(DB$MODALIDADE[DB$TURNO=="Noite" & DB$TEMPO=="1º"])
```

- Híbrido - 5
- Presencial - 6
- **Remoto - 16**

2. Escolha dos alunos que estudam no último tempo no turno da noite.

```
#table(DB$MODALIDADE[DB$TURNO=="Noite" & DB$TEMPO=="3º"])
```

- Híbrido - 16
- Presencial - 5
- **Remoto - 43**

3. Escolha dos alunos que estudam no primeiro tempo do turno da manhã.

```
#table(DB$MODALIDADE[DB$TURNO=="Manhã" & DB$TEMPO=="1º"])
```

- Híbrido - 5
- Presencial - 4
- **Remoto - 9**

4. Escolha dos alunos que estudam no último tempo no turno da manhã.

```
#table(DB$MODALIDADE[DB$TURNO=="Manhã" & DB$TEMPO=="3º"])
```

- Híbrido - 21
- Presencial - 17
- **Remoto - 41**

5. Quantidade de Homens e Mulheres separado por Turma

```
#table(DB$TURMA,DB$SEXO)
```

(Feminino | Masculino)

- T123 - 9 | **36**
- T126 - 4 | **13**
- T132 - 1 | **36**
- T145 - 2 | **16**
- T146 - 3 | **19**
- T147 - 4 | **33**
- T161 - 3 | **24**
- T163 - 1 | **18**
- T927 - 10 | **32**
- T962 - 2 | **8**

Esse resultado, demonstra que em todas as turmas que votaram, a maior parte dos alunos são do sexo masculino.

6. Modalidade preferida das alunas, separadas por tempo de aula, independente do turno.

```
#table(DB$MODALIDADE[DB$SEXO=="Feminino"],DB$TEMPO[DB$SEXO=="Feminino"])
```

(Híbrido | Presencial | Remoto)

- 1º - 1 | 1 | 3
- 2º - 8 | 2 | 10
- 3º - 8 | 3 | 13

ANÁLISE ADICIONAL FOCADAS NOS PARES

Após analisarmos cada coluna individualmente, vamos analisar os votos totais relacionados com outra coluna e tirar uma pontuação geral entre eles, para podermos obter um valor médio de sucesso para cada modalidade.

Para esse experimento as seguintes regras serão utilizadas:

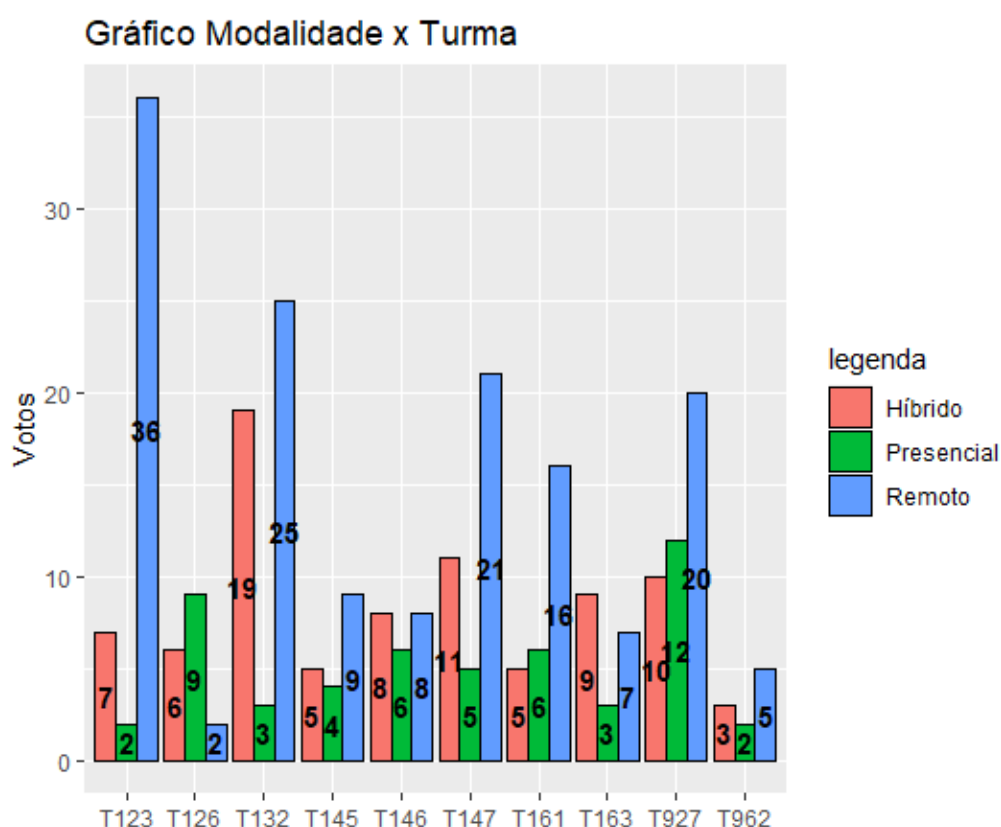
- Para cada **1º lugar** obtido, a modalidade ganha 3 pontos;
- Para cada **2º lugar** obtido, a modalidade ganha 2 pontos;
- Para cada **3º lugar** obtido, a modalidade ganha apenas 1 ponto.
- Caso aconteça um **empate**, as duas modalidades irão ganhar a mesma pontuação.

Através desse experimento, pretende-se tirar uma pontuação geral para cada modalidade, e exemplificar de uma forma alternativa, qual teve mais preferência através das análises dos pares Coluna X Modalidade.

ANÁLISES FOCADAS NO PAR MODALIDADE X TURMA

Este gráfico dispõe das informações sobre a escolha das modalidades de aula com base nas turmas. A modalidade presencial só esteve em maioria na T126, classe de algoritmos II, na segunda-feira, no 2º tempo.

Na turma T163, de algoritmos II, do terceiro tempo teve a modalidade híbrida superior às demais modalidades e a turma T146, teve a modalidade híbrida empatada com a remota, nas demais turmas a modalidade remota foi superior, em 7 turmas, das 10 turmas, a modalidade remota foi superior.



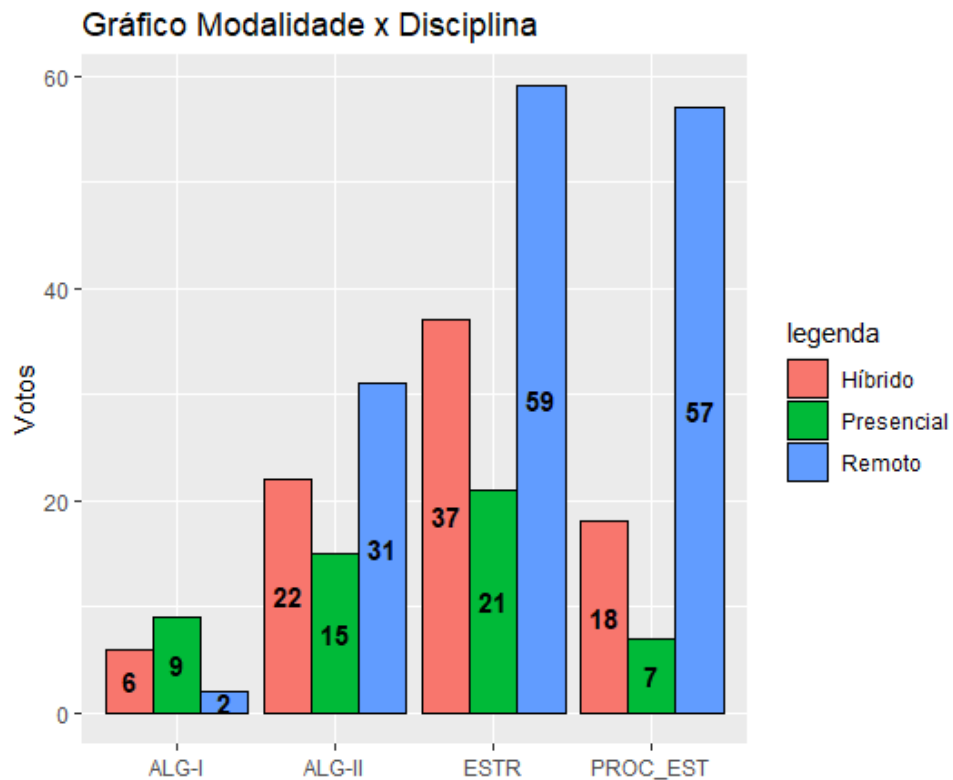
Resultado total:

- Híbrido - (2 | 6 | 2) = 20 Pontos
- Presencial - (1 | 2 | 7) = 17 Pontos
- Remoto - (8 | 1 | 1) = **27 Pontos**

Caso as turmas fossem criadas com base nessas estatísticas, vemos que apenas as turmas 126, 146 e 163 iriam preferir que as suas modalidades não fossem apenas remotas e sim, respectivamente, Presencial, Híbrido ou Remota e por último, Híbrido.

ANÁLISES FOCADAS NO PAR MODALIDADE X DISCIPLINA

Olhando para o comportamento desse gráfico, todas as disciplinas se mantêm na mesma proporção (Remoto > Híbrido > Presencial) exceto para turma de Algoritmos 1, já que a modalidade presencial acaba tendo a maioria dos votos e a modalidade remota a minoria.



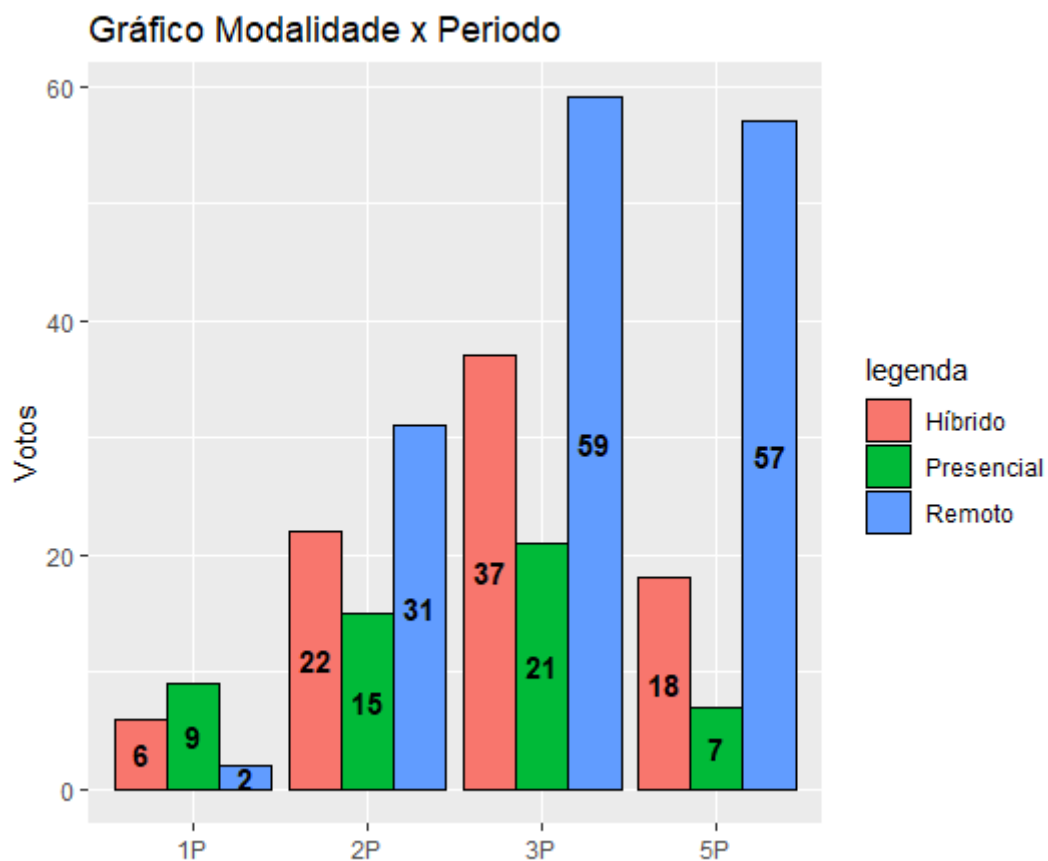
Resultado total:

- Híbrido - (0 | 4 | 0) = 8 Pontos
- Presencial - (1 | 0 | 3) = 6 Pontos
- Remoto - (3 | 0 | 1) = **10 Pontos**

Caso as turmas fossem criadas com base nessas estatísticas, vemos que apenas a matéria de Algoritmos 1 preferiu a modalidade presencial em vez da remota para o seu curso, embora poucos votos dessa turma tenham sido obtidos.

ANÁLISES FOCADAS NO PAR MODALIDADE X PERÍODO

Da mesma forma que os dados relacionados à Disciplina, o Período se mantém sem alteração, visto que os dados são exatamente os mesmos, as proporções se mantêm exceto no primeiro período, com a vitória da modalidade presencial.



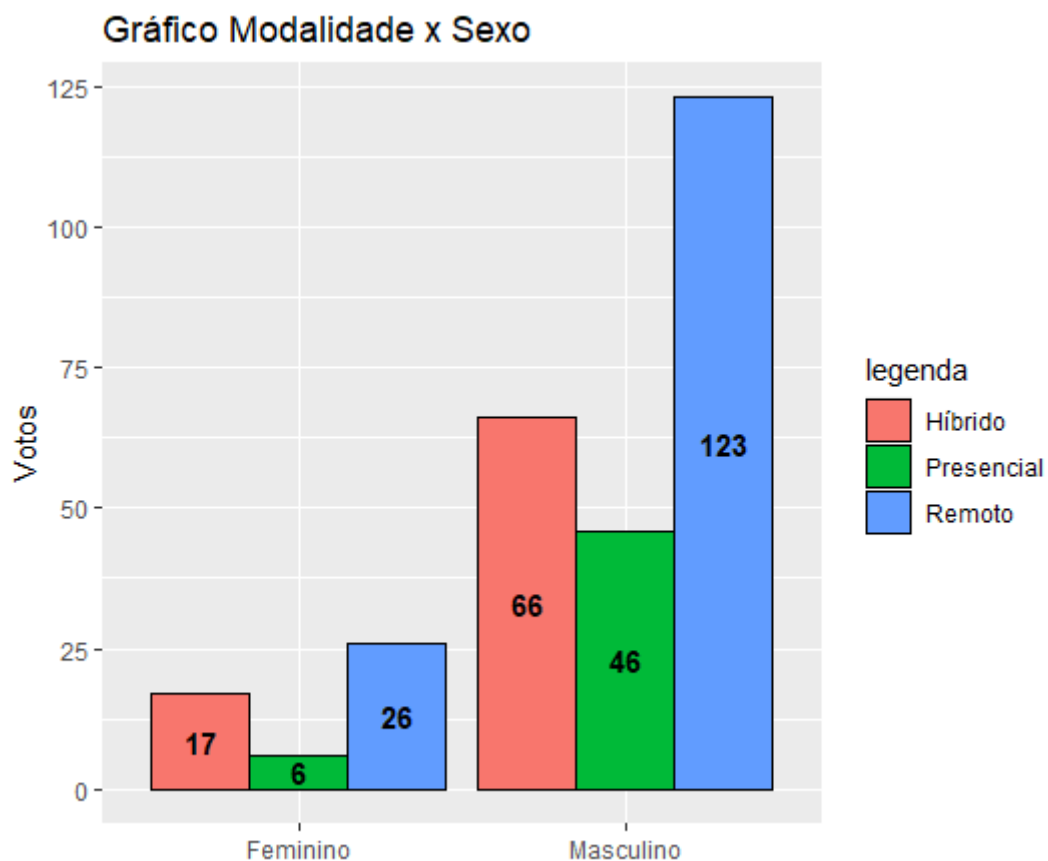
Resultado total:

- Híbrido - (0 | 4 | 0) = 8 Pontos
- Presencial - (1 | 0 | 3) = 6 Pontos
- Remoto - (3 | 0 | 1) = **10 Pontos**

Caso as turmas fossem criadas com base nessas estatísticas, vemos que apenas os alunos do primeiro período preferiram a modalidade presencial em vez da remota para o seu curso, embora poucos votos dessa turma tenham sido obtidos.

ANÁLISES FOCADAS NO PAR MODALIDADE X TURMA

No gráfico abaixo é mostrado que a preferência de ambos os gêneros é pela modalidade remota, se mantendo na mesma proporção, tendo maior diferença por causa da quantidade de votos e não das opções escolhidas, sendo que, a opção presencial foi a menos votada.



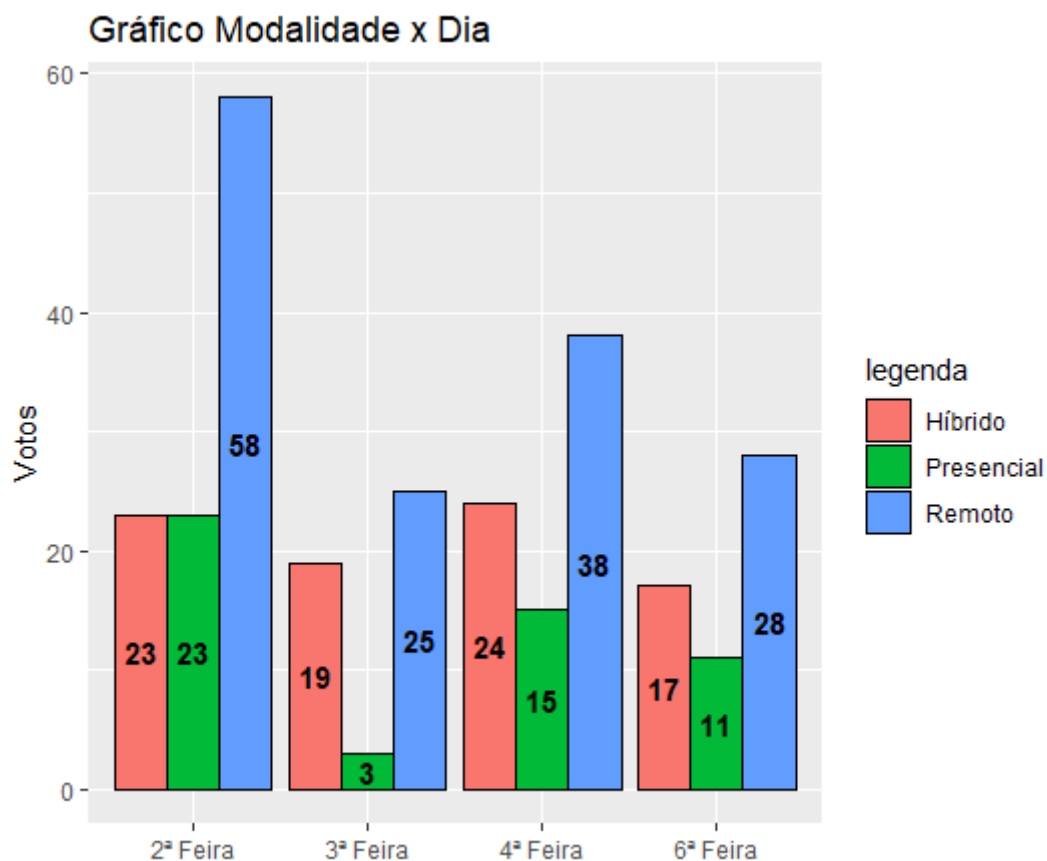
Resultado total:

- Híbrido - (0 | 2 | 0) = 4 Pontos
- Presencial - (0 | 0 | 2) = 2 Pontos
- Remoto - (2 | 0 | 0) = **6 Pontos**

Caso as turmas fossem criadas com base nessas estatísticas, vemos que independente do sexo, a preferência geral se mantém pela modalidade remota, e que as proporções dos votos se mantem tanto no sexo masculino quanto no sexo feminino, repetindo o mesmo comportamento em escala menor visto a pequena quantidade de alunas do sexo feminino.

ANÁLISES FOCADAS NO PAR MODALIDADE X DIA

A modalidade remota foi eleita por alunos de todos os dias da semana, é possível notar que na segunda-feira houve uma grande diferença na escolha da modalidade remota comparada com as outras modalidades.



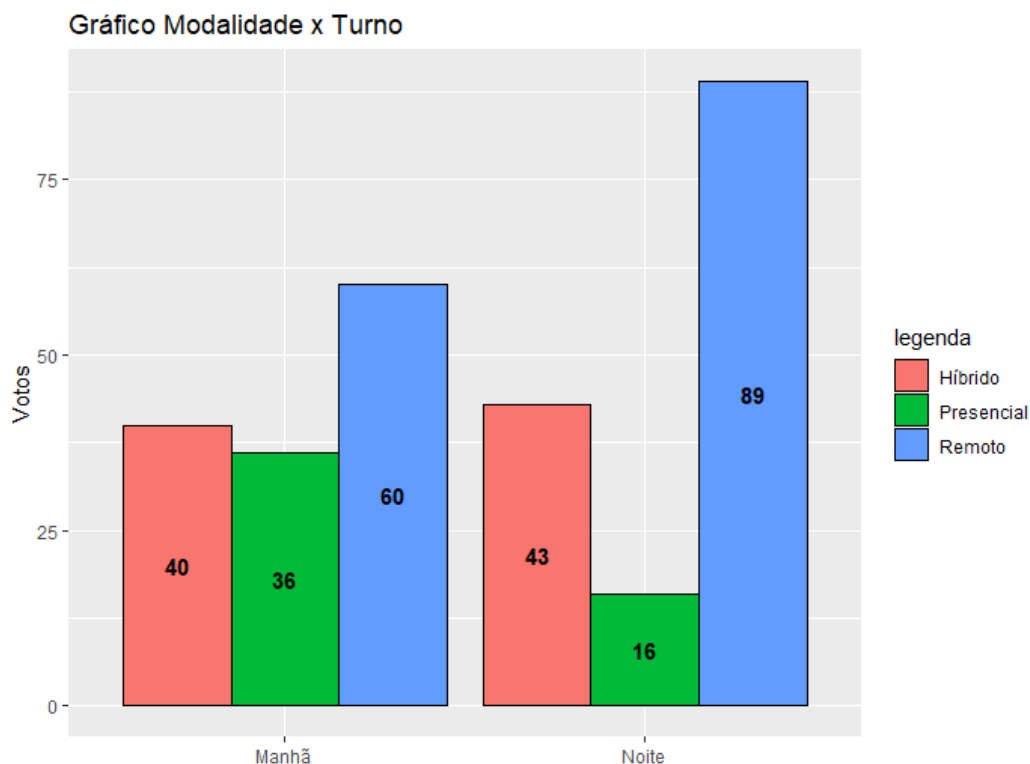
Resultado total:

- Híbrido - (0 | 4 | 0) = 8 Pontos
- Presencial - (0 | 1 | 3) = 5 Pontos
- Remoto - (4 | 0 | 0) = **12 Pontos**

Caso as turmas fossem criadas com base nessas estatísticas, vemos que a modalidade remota segue na liderança, independente do dia da semana, entretanto possui forte preferência no dia de segunda-feira em relação aos outros dias da semana.

ANÁLISES FOCADAS NO PAR MODALIDADE X TURNO

Agora em relação aos turnos (Manhã e Noite) a escolha da maioria foi pela modalidade remota, entretanto, pode-se observar uma ocorrência interessante, no turno da noite fica muito mais acentuada a escolha por essa modalidade.



Resultado total:

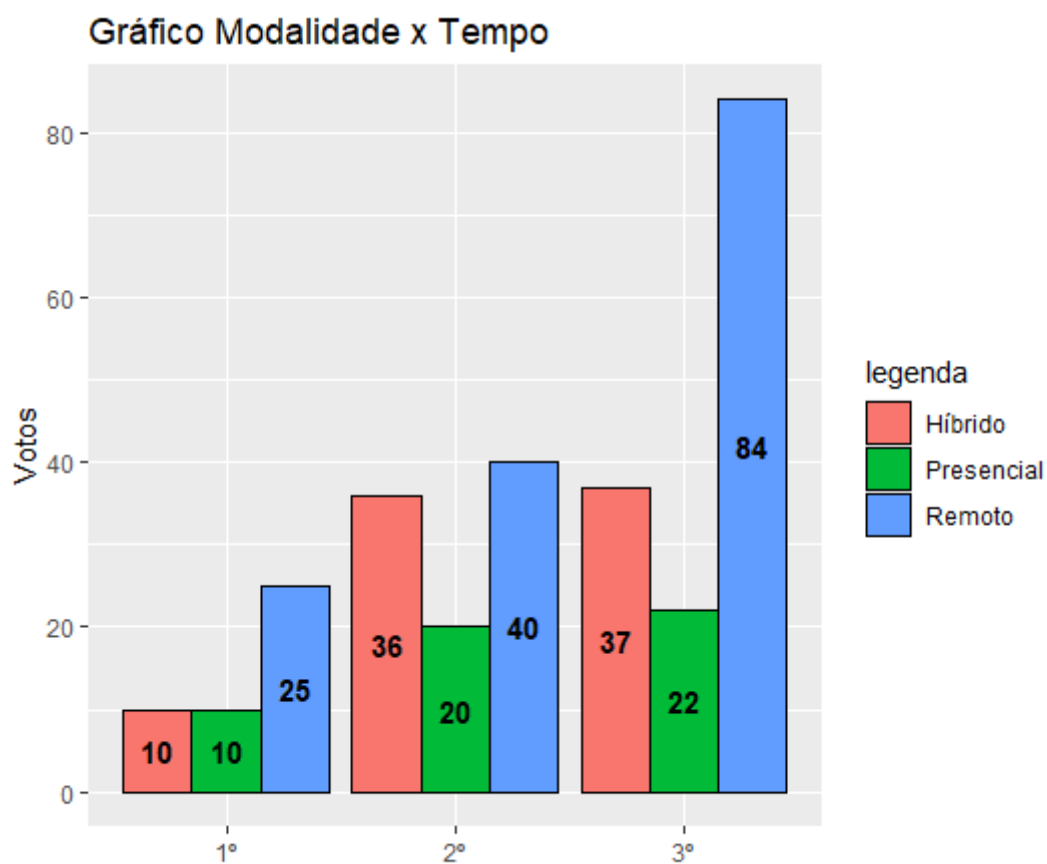
- Híbrido - (0 | 2 | 0) = 4 Pontos
- Presencial - (0 | 0 | 2) = 2 Pontos
- Remoto - (2 | 0 | 0) = **6 Pontos**

Caso as turmas fossem criadas com base nessas estatísticas, podemos observar que a modalidade remota também segue ganhando porém com um ponto interessante a se observar, no turno da noite, a preferência por aulas remotas é muito maior do que a preferência pela modalidade presencial, por diversos possíveis motivos, como por exemplo, o perigo de andar nas ruas por volta da noite.

Já no turno da manhã, a diferença é relativamente bem menor, mas ainda sim existe um salto entre as modalidades presencial e remota.

ANÁLISES FOCADAS NO PAR MODALIDADE X TEMPO

Observando o gráfico de Modalidade x Tempo, pode-se perceber que quanto mais tarde for o tempo em que os alunos estão estudando, maior é a vontade geral pela modalidade remota.



Resultado total:

- Híbrido - (0 | 3 | 0) = 6 Pontos
- Presencial - (0 | 1 | 2) = 4 Pontos
- Remoto - (3 | 0 | 0) = **9 Pontos**

Caso as turmas fossem criadas com base nessas estatísticas, pode-se observar que, quanto mais tarde for o tempo que o aluno estuda, maior é a vontade de que a aula dele seja da modalidade remota, uma quantidade enorme de alunos do terceiro tempo preferem a modalidade remota, maior do que a soma de sua parte híbrida e presencial.

PONTUAÇÕES DA ANÁLISE ADICIONAL:

Turma:

Híbrido - (2 | 6 | 2)

Presencial - (1 | 2 | 7)

Remoto - (8 | 1 | 1)

Dia:

Híbrido - (0 | 4 | 0)

Presencial - (0 | 1 | 3)

Remoto - (4 | 0 | 0)

Disciplina:

Híbrido - (0 | 4 | 0)

Presencial - (1 | 0 | 3)

Remoto - (3 | 0 | 1)

Turno:

Híbrido - (0 | 2 | 0)

Presencial - (0 | 0 | 2)

Remoto - (2 | 0 | 0)

Período:

Híbrido - (0 | 4 | 0)

Presencial - (1 | 0 | 3)

Remoto - (3 | 0 | 1)

Tempo:

Híbrido - (0 | 3 | 0)

Presencial - (0 | 1 | 2)

Remoto - (3 | 0 | 0)

Turma:

Híbrido - (0 | 2 | 0)

Presencial - (0 | 0 | 2)

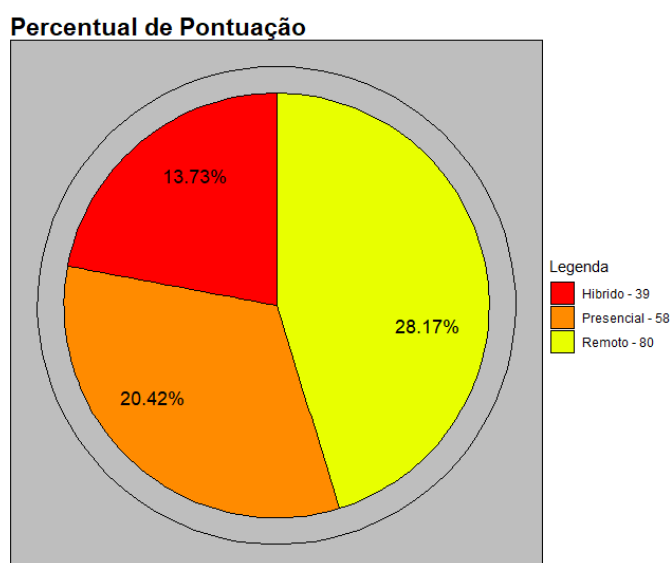
Remoto - (2 | 0 | 0)

PONTUAÇÕES TOTAIS:

Híbrido - (2 | 25 | 2) = 58 Pontos

Presencial - (3 | 4 | 22) = 39 Pontos

Remoto - (25 | 1 | 3) = **80 Pontos**



Com isso, podemos observar ainda que a modalidade mais desejada de forma geral pela maioria dos alunos, independente de suas características, é a modalidade **Remota** e a modalidade menos desejada é a modalidade **Presencial**

ANÁLISE DE MINERAÇÃO DE DADOS COM REGRAS DE ASSOCIAÇÃO

Foi feita a mineração de regras de associação entre os dados do documento, onde cada linha constitui uma transação, ou o registro atômico (individual) de vários dados identificadores. Desta forma, é possível identificar a probabilidade de um conjunto de dados A condicionar a aparência de um conjunto de dados B.

Os parâmetros de mensuração das transações são:

Antecedente: Conjunto de itens A condicionador de um conjunto consequente B

Consequente: Conjunto de itens B ocasionado pelo conjunto antecedente

Confiança esperada: trata-se da probabilidade de ocorrência do conjunto consequente

As métricas para aferir o condicionamento de conjuntos são

Suporte: Quantidade de itens em um conjunto dividido pela quantidade total de itens de todas as transações

Confiança: Probabilidade condicional de um conjunto de itens A implicar na aparição de um conjunto de itens B. É dado pela razão da probabilidade da intersecção entre antecedente e consequente dividido pela probabilidade do antecedente.

Alavancagem(Lift): É a razão entre a confiança de um conjunto A implicar num conjunto B e a probabilidade de ocorrência de B (conhecido como confiança esperada do consequente).

Ao executar o apriori sobre os dados tratados do documento pesquisa.xlsx, obtemos os seguintes resultados:

```
Parameter specification:
confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext
0.1 0.1 1 none FALSE TRUE 5 0.05 1 10 rules TRUE

Algorithmic control:
filter tree heap memopt load sort verbose
0.1 TRUE TRUE FALSE TRUE 2 TRUE

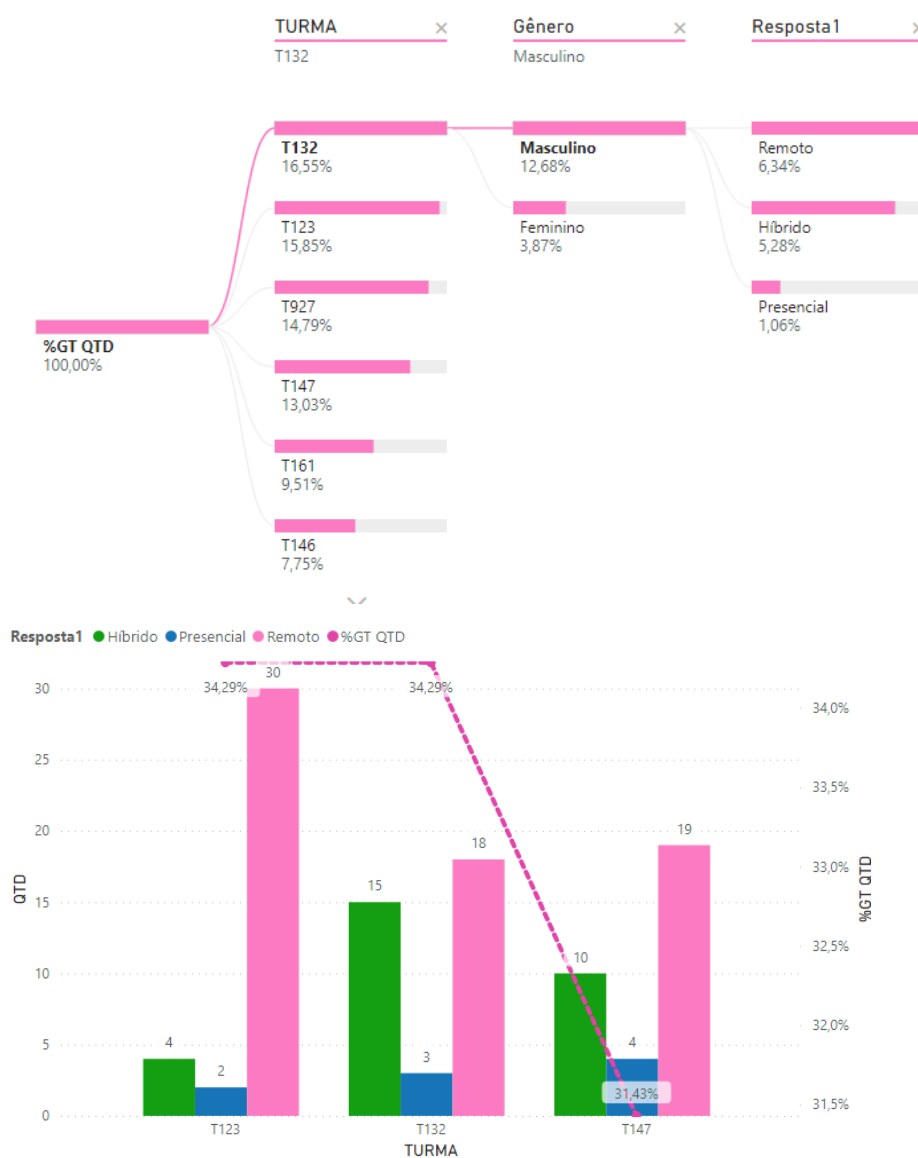
Absolute minimum support count: 14

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[32 item(s), 284 transaction(s)] done [0.00s].
sorting and recoding items ... [16 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing ... [17 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
> dfdatarulesdf=data.frame(lhs=labels(lhs(dfdatarules)),rhs=labels(rhs(dfdatarules)),dfdatarules@quality)
> dfdatarulesdf
```

Foram especificados os parâmetros de 5% de suporte, 10% de confiança, conjuntos com tamanho mínimo de 1 item e máximo de 10 itens.

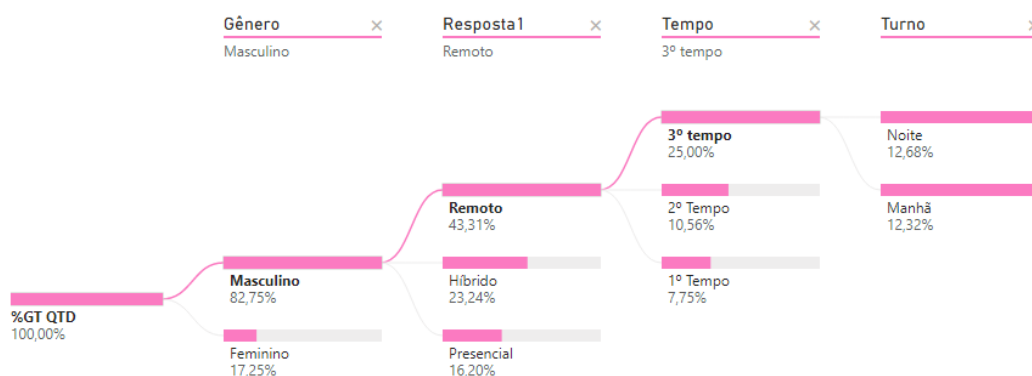
	lhs	rhs	support	confidence	coverage	lift	count
1	{}	{,"ESTR","3P",1,"H","H"}	0.10211268	0.1021127	1.0000000	1.000000	29
2	{}	{T147}	0.13028169	0.1302817	1.0000000	1.000000	37
3	{}	{T927}	0.14788732	0.1478873	1.0000000	1.000000	42
4	{}	{,"ESTR","3P",1,"R","R"}	0.15845070	0.1584507	1.0000000	1.000000	45
5	{}	{T123}	0.15845070	0.1584507	1.0000000	1.000000	45
6	{}	{T132}	0.16549296	0.1654930	1.0000000	1.000000	47
7	{}	{,"PROC_EST","5P",1,"R","R"}	0.17253521	0.1725352	1.0000000	1.000000	49
8	{,"ESTR","3P",1,"H","H"}	{T132}	0.05281690	0.5172414	0.1021127	3.125459	15
9	{T132}	{,"ESTR","3P",1,"H","H"}	0.05281690	0.3191489	0.1654930	3.125459	15
10	{T147}	{,"PROC_EST","5P",1,"R","R"}	0.06690141	0.5135135	0.1302817	2.976282	19
11	{,"PROC_EST","5P",1,"R","R"}	{T147}	0.06690141	0.3877551	0.1725352	2.976282	19
12	{T927}	{,"ESTR","3P",1,"R","R"}	0.05633803	0.3809524	0.1478873	2.404233	16
13	{,"ESTR","3P",1,"R","R"}	{T927}	0.05633803	0.3555556	0.1584507	2.404233	16
14	{,"ESTR","3P",1,"R","R"}	{T132}	0.06338028	0.4000000	0.1584507	2.417021	18
15	{T132}	{,"ESTR","3P",1,"R","R"}	0.06338028	0.3829787	0.1654930	2.417021	18
16	{T123}	{,"PROC_EST","5P",1,"R","R"}	0.10563380	0.6666667	0.1584507	3.863946	30
17	{,"PROC_EST","5P",1,"R","R"}	{T123}	0.10563380	0.6122449	0.1725352	3.863946	30

As regras de associação mais significativas encontradas foram as que associam a turma T132 de estruturas de dados tanto às modalidades Híbrida quanto Remota e votantes do sexo masculino, assim como as turmas 123 e 147 de processos estocásticos também as modalidades Remota e Híbrida e ao mesmo perfil de votantes.



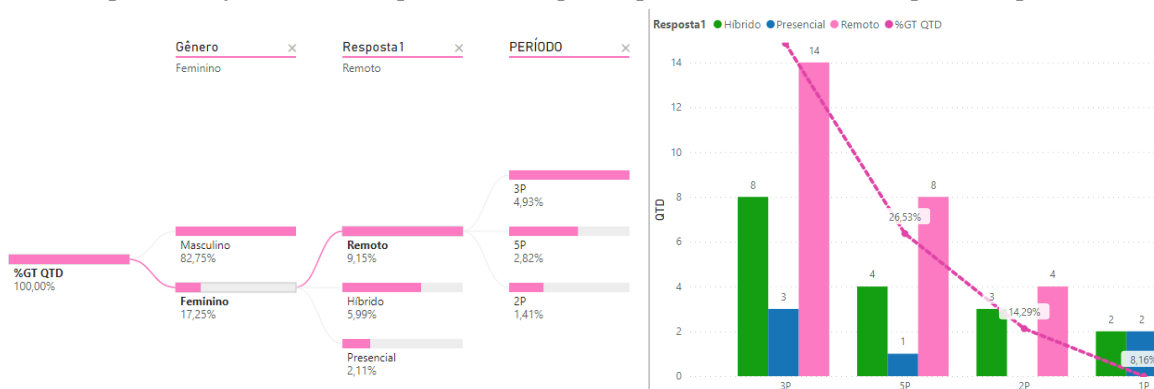
As turmas de final 2 e 3 são correspondentes ao segundo e terceiro tempo da noite, e a de final 7 corresponde ao último tempo da manhã. O segundo número corresponde a data, desta forma compreendem as turmas de terça-feira (3), segunda-feira (2) e quarta-feira (4).

Com base nestas regras de associação, podemos aferir que os últimos tempos da noite e o último tempo da manhã possuem um número maior de alunos do sexo masculino e optam majoritariamente por ensino remoto, em seguida há manifestação considerável de predileção pelo ensino híbrido, ao passo que a opção de ensino presencial é pouco frequente em relação ao analisado, considerando 5% de frequência em qualquer conjunto.



A quantidade de alunos do sexo feminino (2) também é pouco presente nos conjuntos de itens. Os períodos são itens frequentes dos conjuntos, pois são redundâncias a cada disciplina, sendo 3P(3º período) para estruturas de dados e 5P(5º período) para processos estocásticos.

A alavancagem destas regras destacadas é superior a 3, e isso demonstra a relação numérica em que a quantidade de ocorrências do antecessor ocasiona o consequente. Estas regras relacionam o número da turma as características de período, disciplina, sexo e modalidade escolhida, e têm forte relação nos dois sentidos. Ou seja, tanto se pode aferir que turmas específicas das matérias de processos estocásticos e estruturas de dados optam primeiro por ensino remoto e em seguida híbrido e são compostas majoritariamente por homens, quanto pode-se inferir a consequência oposta.



Analisou-se a frequência dos itens, caracterizada pela métrica de suporte, para analisar quais são os conjuntos de maior repetição na lista de transações do dataset. Podemos com isto analisar os conjuntos de itens mais ocorridos, mapeando quais os dados são mais expressivos para a análise de mineração.

```

1/ {"PROC_EST","5P",1,"R","R"} {1123} 0.10563380 0.6122449 0.1725352 3.863946 30
> frequentItems <- eclat (dfdata_tran, parameter = list(supp = 0.05, maxlen = 10))
Eclat

parameter specification:
tidLists support minlen maxlen target ext
FALSE 0.05 1 10 frequent itemsets TRUE

algorithmic control:
sparse sort verbose
7 -2 TRUE

Absolute minimum support count: 14

create itemset ...
set transactions ... [32 item(s), 284 transaction(s)] done [0.00s].
sorting and recoding items ... [16 item(s)] done [0.00s].
creating sparse bit matrix ... [16 row(s), 284 column(s)] done [0.00s].
writing ... [21 set(s)] done [0.00s].
Creating S4 object ... done [0.00s].
> |

```

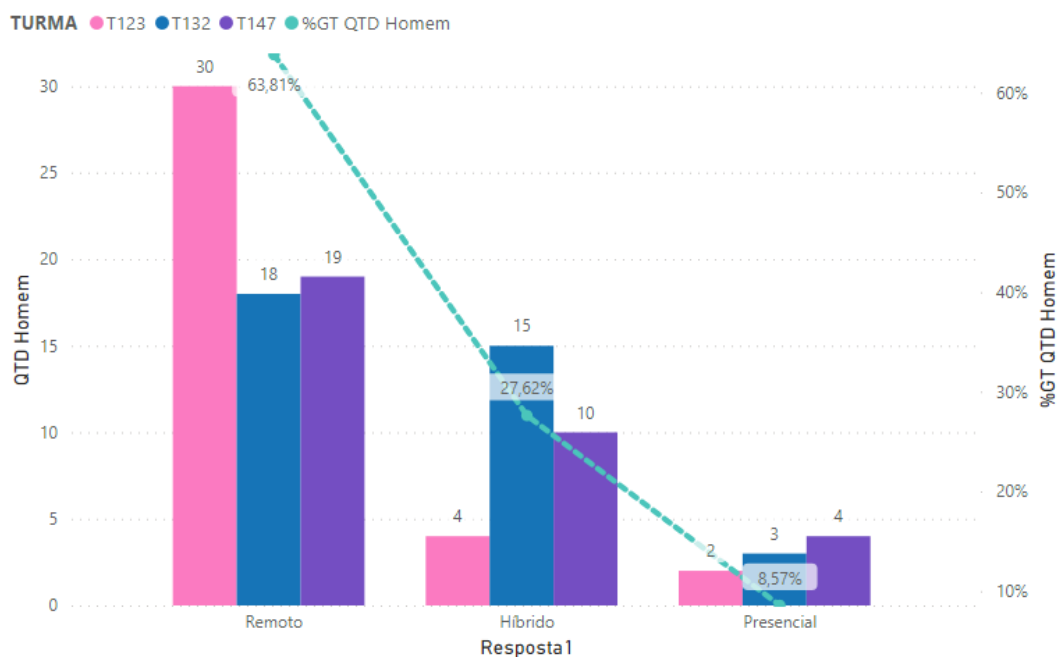
Listagem dos itens mais frequentes com o inspect.

```

Creating S4 object ... done [0.00s].
> inspect(frequentItems)
      items                                support  transIdenticalToItemsets count
[1] {"ESTR","3P",1,"H","H",T132} 0.05281690 15 15
[2] {"PROC_EST","5P",1,"R","R",T147} 0.06690141 19 19
[3] {"ESTR","3P",1,"R","R",T927} 0.05633803 16 16
[4] {"PROC_EST","5P",1,"R","R",T123} 0.10563380 30 30
[5] {"ESTR","3P",1,"R","R",T132} 0.06338028 18 18
[6] {"PROC_EST","5P",1,"R","R"} 0.17253521 49 49
[7] {T132} 0.16549296 47 47
[8] {"ESTR","3P",1,"R","R"} 0.15845070 45 45
[9] {T123} 0.15845070 45 45
[10] {T927} 0.14788732 42 42
[11] {T147} 0.13028169 37 37
[12] {"ESTR","3P",1,"H","H"} 0.10211268 29 29
[13] {T161} 0.09507042 27 27
[14] {"ALG-II","2P",1,"R","R"} 0.09507042 27 27
[15] {T146} 0.07746479 22 22
[16] {"ALG-II","2P",1,"H","H"} 0.06690141 19 19
[17] {T163} 0.06690141 19 19
[18] {"ESTR","3P",1,"P","P"} 0.06338028 18 18
[19] {T145} 0.06338028 18 18
[20] {T126} 0.05985915 17 17
[21] {"ALG-II","2P",1,"P","P"} 0.05281690 15 15
> |

```

Assim como nas regras de associação mineradas, os itens mais frequentes correspondem às disciplinas de estruturas de dados e processos estocásticos, sexo masculino e opção por modalidades remotas ou híbridas e as turmas T123, T132 e T147. Também aparecem outros conjuntos frequentes que contenham as turmas T927 de estrutura de dados do último tempo da manhã na segunda-feira, e a disciplina de algoritmos II com modalidades presenciais e remotas.



Por fim, podemos concluir da mineração que a maior quantidade de votos se concentra em turmas das matérias de processos estocásticos, estruturas de dados e algoritmos II, as modalidades preferidas são remota e híbrida e a maioria dos alunos são do sexo masculino nestas turmas. O turno da noite é onde houve maior frequência de itens, especialmente os dois últimos tempos, e o turno da manhã aparece mais frequentemente no último tempo.

CONCLUSÕES FINAIS

Com todo o levantamento descritivo e a extração de conhecimento das relações entre os dados analisados, podemos aferir as seguintes assertivas com segurança em relação aos alunos que votaram nas duas pesquisas sobre preferência por modalidade de ensino. Foram analisados apenas os votos válidos, **sendo desconsiderados do dataset os alunos que não votaram.**

Mais de 80% dos alunos que votaram na pesquisa são do sexo masculino nas turmas onde foi realizada pesquisa. O sexo masculino é maior parte da composição da amostragem de turmas, e reflete o resultado obtido pelos votos dispostos.

A maior parte dos votos foram coletados da turma de Estrutura de dados (Terceiro período), tendo uma porcentagem total de **41.2%** dos votos, em seguida a turma de Processos Estocásticos com **28.87%** e Algoritmos II com **23.04%**. A turma de Algoritmos I compôs menos de 6% dos votos.

A maior parte dos votos foram obtidos de alunos que estudavam em uma segunda-feira, sendo um total de 104 votos, ou seja, um percentual de **36.62%** dos votos totais coletados. Os votos dos alunos que estudam na quarta-feira estão em segundo lugar com **27.11%**, e na sexta-feira com **19.72%**. A maior parte dos votantes estuda na terça-feira ou na quarta-feira.

Ao analisar o tempo de aula na amostragem da pesquisa, percebe-se que **quanto mais tarde o tempo maior foi a quantidade de votos**, sendo os alunos do 3º tempo os que mais tiveram influência no resultado da votação com 50.35% de votos e os que menos tiveram influência foram os alunos do 1º tempo com 15.85%. Os alunos de 2º tempo compõe 33.8% dos votos.

Houve pouca diferença entre o turno das disciplinas na composição de votos, sendo mais votantes à noite com **52.11%** do total. Isso demonstra a neutralidade da amostra que, ainda que tenha mais integrantes do turno noturno, está equilibrada em quantidade similar para ambos os turnos, sem que um seja mais representativo que outro.

Referente às turmas, a distribuição de votos teve pouco diferença entre elas, sendo 4 turmas as que compõem mais de 10% dos votos cada: 132 (**16.55%**) 123 (**15.85%**) 927 (**14.79%**) e 147 (**13.03%**), o que implica em maior peso nos votos totais. A turma 962 teve menos de 10% dos votos (**3.52%**), por isso exerce baixa influência no total. As demais turmas contribuem de forma equilibrada para os votos.

Em relação ao período dos alunos que responderam a pesquisa, **o terceiro período (3P) é o que concentrou mais votantes**, compondo **41.2%** dos votos totais. Os outros períodos com mais votos foram o 5º período (5P) com 28.87% e o 2º período (2P) com 23.94% dos votos. O 1º período (1P) caracteriza aproximadamente 6% do total, por isto pouco representa toda a amostragem. Esta informação corrobora as disciplinas com maior quantidade de votos, que são estruturas de dados (3P) e processos estocásticos (5P), e em seguida algoritmos II (2P).

Entre os votantes, **a modalidade de ensino com maior preferência é o remoto** com um total de 149 votos (**52.46%**), ou seja **mais do que a soma de seus concorrentes**, híbrido com 83% e presencial com 52%. Logo **em seguida, a modalidade com mais votos é o modelo Híbrido**, com 29.13%, já o presencial possui 18.31% do total. Juntos, **os modelos híbrido e presencial totalizam 47,55% do total**. Portanto, dentre os votos obtidos, o modelo presencial tem baixíssima predileção, ao passo que o remoto é o vitorioso e o modelo híbrido tem uma preferência expressiva. Com isto, podemos concluir que **há baixa adesão ao regime integralmente presencial** por parte dos alunos que votaram, e **os modelos remoto e híbrido são os mais requisitados**, especialmente **o modelo totalmente remoto**.

A amostra é mais frequentemente caracterizada por alunos das disciplinas de Estruturas de dados e Processos estocásticos, com uma quantidade significativa de alunos de algoritmos II, e poucos alunos da disciplina de algoritmos I. A maioria esmagadora dos votantes é do sexo masculino. Os tempos de aula mais frequentes são o último tempo e o segundo, sendo poucos votos de primeiro tempo. O terceiro período concentra mais votos, e logo depois o 5º período. Os dias da semana com maior quantidade de votos são, respectivamente, segunda, quarta e sexta-feira. O turno da noite e o da manhã possuem paridade de 52% para 48% aproximadamente. Houve pouca diferença de voto entre cada turma, sendo a com menos votos a turma 962, com apenas 10 votos, todas as demais com mais de 15 votos e as turmas maiores com mais de 40.

Por fim, é possível afirmar assertivamente que **turmas de Estruturas de Dados e Processos Estocásticos** tanto da manhã quanto da noite possuem **a maior parte dos alunos de sexo masculino, optam massivamente pelo ensino remoto e em grande parte pelo híbrido, e têm baixa preferência pelo ensino totalmente presencial**.

REFERÊNCIAS BIBLIOGRÁFICAS

<https://www.datasciencemadesimple.com/delete-or-drop-rows-in-r-with-conditions-2/>

<https://www.guru99.com/r-data-frames.html#:~:text=We%20can%20create%20a%20dataframe,the%20name%20of%20the%20variables.>

https://docs.oracle.com/cd/E18283_01/datamine.112/e16808/algo_apriori.htm#:~:text=generate%20fewer%20rules-,Metrics%20for%20Association%20Rules,supports%20lift%20for%20association%20rules.

<http://r-statistics.co/Association-Mining-With-R.html>

<https://www.rdocumentation.org/packages/arules/versions/1.7-3/topics/apriori>

<https://towardsdatascience.com/association-rules-2-aa9a77241654>