

Análise Cursos UFCG

Caynan Sousa

17/11/2016

Nossos Dados

Iremos analisar os dados do controle acadêmico da UFCG, nossos dados contemplam os períodos de 2002 a 2014 e todos os cursos da UFCG.

Esses dados estão disponíveis em um **dataframe** chamado **alunosUFCG**. Podemos ver então que os atributos disponíveis, são:

```
colnames(alunosUFCG)
```

```
## [1] "Matricula"      "Cod_Curso"      "Cod_Disciplina"
## [4] "Cod_Evasao"     "Nome_Curso"     "Periodo"
## [7] "Nome_Disciplina" "Creditos"       "Departamento"
## [10] "Media_Disciplina" "Situacao"       "Centro"
## [13] "Campus"         "Creditos_Curso" "Tipo"
## [16] "Periodo_Ingresso" "Periodo_Relativo"
```

Como podemos ver todos os dados que temos dos alunos é bem auto-explicativo, a parter, talvez dos seguintes campos:

- **Cod_Evasao:** 0 se o aluno *não* evadiu ao fim do primeiro periodo e 1 se ele evadiu.
- **Credito_Curso:** Total de créditos necessários para o aluno concluir o curso.
- **Periodo_Ingresso:** período que o aluno entrou no curso.

Perguntas

Separamos então algumas perguntas interessantes que podemos responder com ajuda desses dados, além de uma previsão inicial baseada totalmente em intuição, afim de podermos ver se os dados concordam com ~~minha~~ nossa intuição.

1. Qual campus possui a maior quantidade de evasão? E a menor?

Eu imagino que os campus com menos concentração populacional como Cuité, Sumé e Pombal estejam entre os com maior evasão, e simetricamente os com maior concentração populacional, como Campina Grande, Cajazeiras e Patos estejam entre os com menor evasão.

2. Qual disciplina possui a maior quantidade de alunos reprovados por falta?

Eu imagino que provavelmente alguma disciplina do CCT, tais como as que são oferecidas pelos departamentos de matemática e/ou estatística.

3. Existe alguma ligação (correlação) entre ser admitido na universidade no segundo semestre e reprovações?

Eu imagino que sim e que seja uma correlação positiva, ou seja existem mais reprovações entre os alunos admitidos no segundo semestre do que entre os alunos admitidos no primeiro semestre.

Respondendo Nossas Perguntas

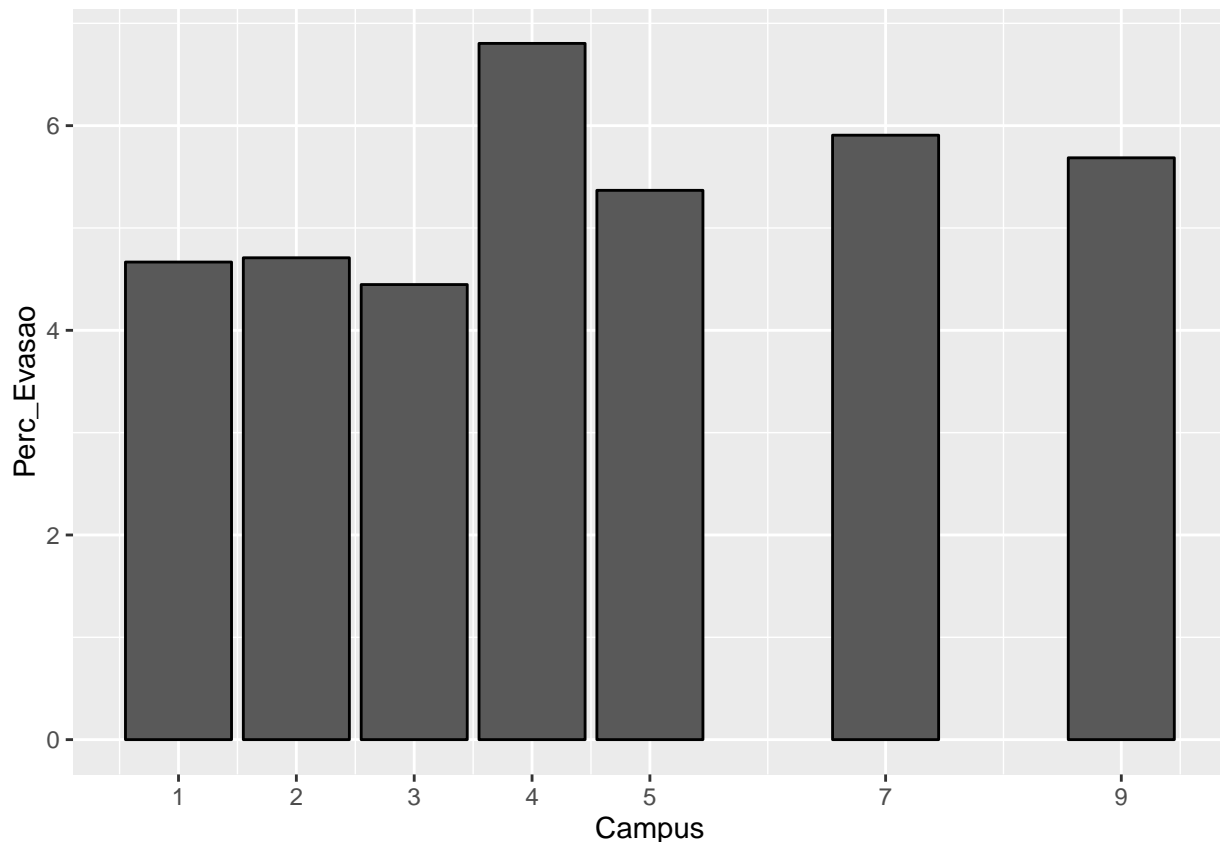
Qual campus possui a maior quantidade de evasão? E a menor?

Para responder essa pergunta, iremos contar a quantidade de evasão em cada campus, e afim de obtermos um resultado normalizado para a quantidade de alunos em cada campus, iremos dividir pelo número de alunos em cada respectivo campus, assim obtendo uma porcentagem de alunos que evadiram a universidade.

```
alunosUFCG.evasao <- filter(alunosUFCG, Cod_Evasao == 1)

evasaoUFCG <- alunosUFCG.evasao %>% filter(!is.na(Cod_Evasao)) %>%
  group_by(Campus) %>%
  summarise(
    Num_Alunos = n_distinct(Matricula),
    Num_Evasoes = n(),
    Perc_Evasao = Num_Evasoes / Num_Alunos
  )
```

Visualmente temos:



Podemos então observar que os campus 4 e 7 são os com maior porcentagem de evasão dentre todos os campus. E que os campus 3 e 1 são os com menor porcentagem de evasão entre os campus observados.

Qual disciplina possui a maior quantidade de alunos reprovados por falta?

Para responder essa pergunta, iremos contar a quantidade de reprovações por falta para cada disciplina durante todo o período observado, e afim de obtermos um resultado normalizado para a quantidade de alunos matriculados em cada disciplina, iremos dividir pelo número de alunos matriculados em cada respectiva disciplina, assim obtendo uma porcentagem dos alunos que foram reprovados por falta para cada disciplina.

```
alunosUFCG.reprovado.falta <- filter(alunosUFCG, Situacao == "Reprovado por Falta")

reprovados.falta <- alunosUFCG.reprovado.falta %>% filter(!is.na(Situacao)) %>%
  group_by(Cod_Disciplina) %>%
```

```

summarise(
  Nome_Disciplina = first(Nome_Disciplina),
  Departamento = first(Departamento),
  Num_Reprovados_Falta = n()
)

todos.alunos <- alunosUFCG %>% filter(!is.na(Situacao)) %>%
  group_by(Cod_Disciplina) %>%
  summarise(
    Num_Alunos = n()
  ) %>%
  arrange(desc(Num_Alunos))

summary.reprovados <- merge(reprovados.falta, todos.alunos, by="Cod_Disciplina") %>%
  mutate(Perc_Reprovados = Num_Reprovados_Falta / Num_Alunos) %>%
  # Removemos disciplinas onde todos os alunos que já a cursaram a perderam por falta,
  # provavelmente são outliers.
  filter(Num_Reprovados_Falta != Num_Alunos & Num_Alunos > 30) %>%
  arrange(desc(Perc_Reprovados))

head(summary.reprovados)

```

```

##   Cod_Disciplina      Nome_Disciplina
## 1      2102250      FÍSICA PARA CIÊNCIAS BIOLÓGICAS
## 2      1108103 PRAT. DE ENS. DA FÍSICA NA EDUC. BASICA
## 3      1108102 TÓPICOS DE HISTÓRIA E ENSINO DA FÍSICA
## 4      7102257      TRABALHO DE CONCLUSAO DE CURSO
## 5      1305075      DIREITO ADMINISTRATIVO
## 6      5101082      CALCULO DIFERENCIAL E INTEGRAL I
##                                Departamento Num_Reprovados_Falta
## 1 UNID. ACAD. DE CIÊNCIAS EXATAS E DA NATUREZA      105
## 2                                UNID. ACAD. DE FÍSICA      15
## 3                                UNID. ACAD. DE FÍSICA      29
## 4 UNID. ACAD. DE TECNOLOGIA DO DESENVOLVIMENTO      24
## 5      UNID. ACAD. DE SOCIOLOGIA E ANTROPOLOGIA      21
## 6                                UNID. ACAD. DE EDUCAÇÃO      467
##   Num_Alunos Perc_Reprovados
## 1         180      0.5833333
## 2          33      0.4545455
## 3          71      0.4084507
## 4          59      0.4067797
## 5          52      0.4038462
## 6        1201      0.3888426

```

O resultado me foi um pouco inesperado, mesmo após reduzir os resultados para disciplinas com 30+ alunos matriculados, principalmente a disciplina na quinta posição, por ser uma disciplina do departamento de Antropologia e Sociologia, que não era um dos departamentos que previ que estaria entre o top 5.