

Data Analyst - Desafio

Análise Exploratória de Dados

Luciano O. Batista

Como melhorar a experiência de aprendizagem dos estudantes?

29/08/2020

Base de Dados

- Foram utilizados dados reais de mais de 30 mil estudantes da Open University (OU). Universidade de ensino online no Reino Unido.
- Os dados estão distribuídos em 7 tabelas, contendo informações sobre:
 - Cursos
 - Estudantes
 - Plataforma online
- Toda informação foi coletada e anonimizada para preservar a privacidade dos estudantes.

Objetivo

- Apresentar insights importantes que possam auxiliar na tomada de decisão e consequentemente melhorar a experiência de aprendizagem dos estudantes.

Público Alvo

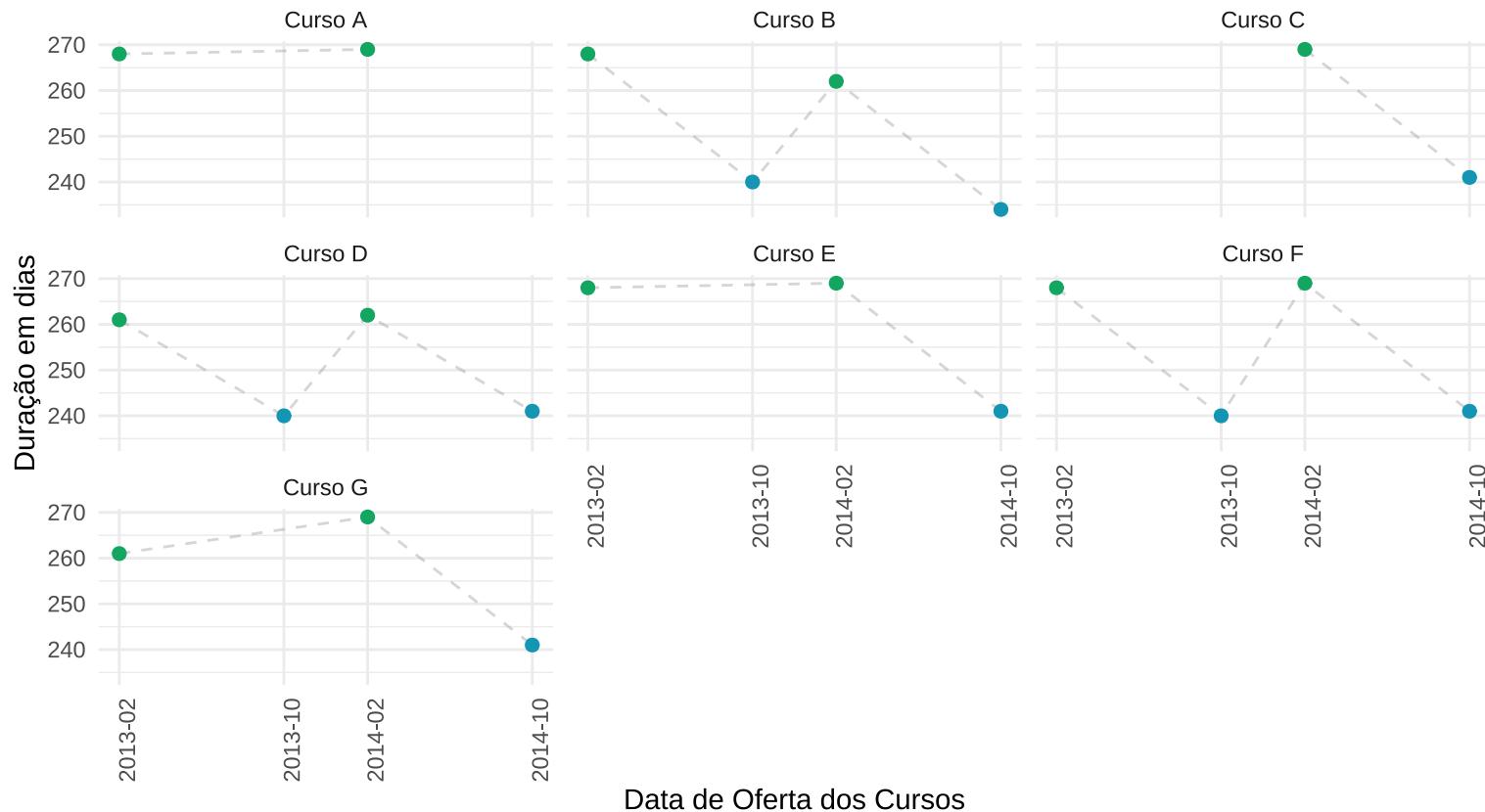
- Gerência

Cursos e Plataforma Online

[*] Alguns cursos não são ofertados por todo período disponível no dataset, e alguns gráficos terão espaços em branco devido a esse aspecto

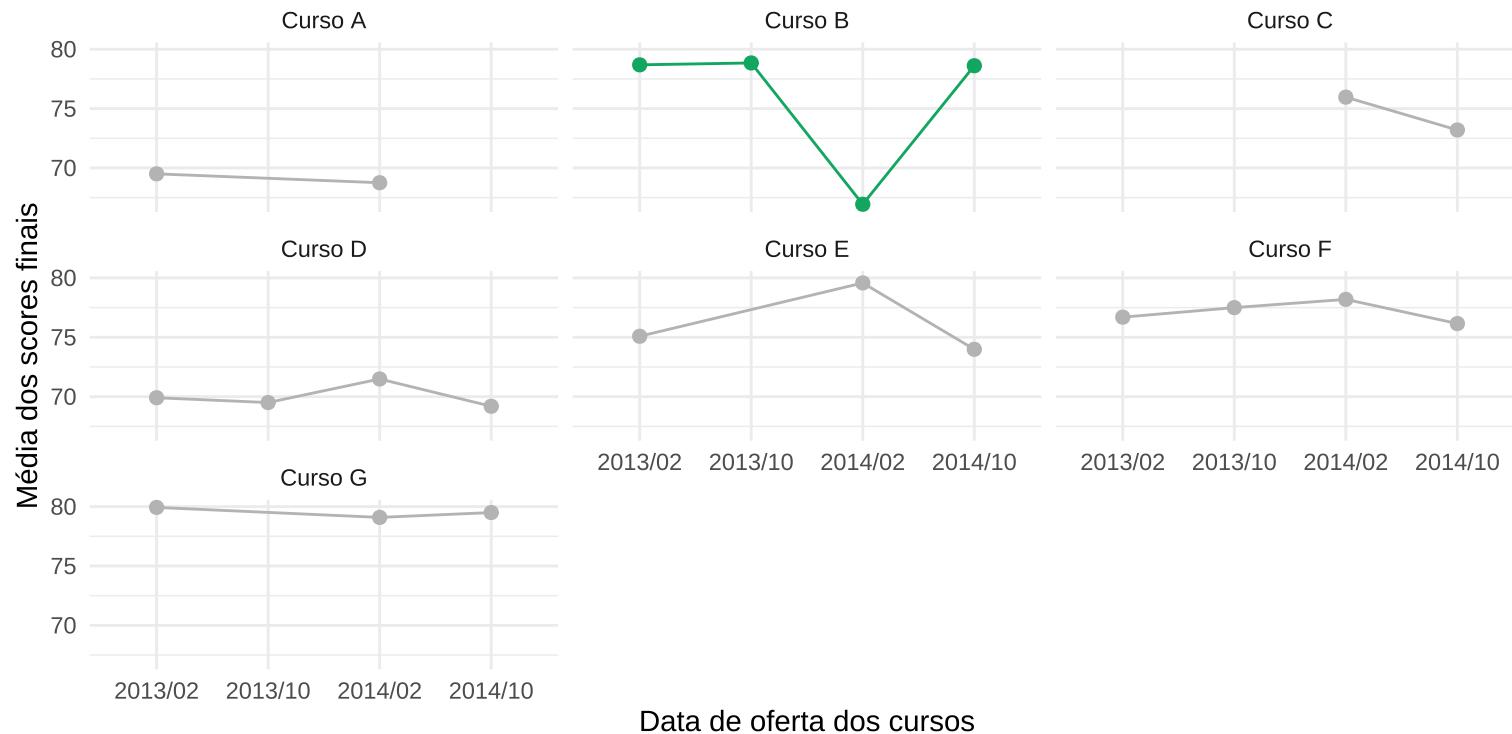
Oferta e Duração dos Cursos

Observou-se que sempre que um curso é oferecido no **primeiro período** do ano ele possui menor carga horária total (em dias), comparado ao **segundo período**.



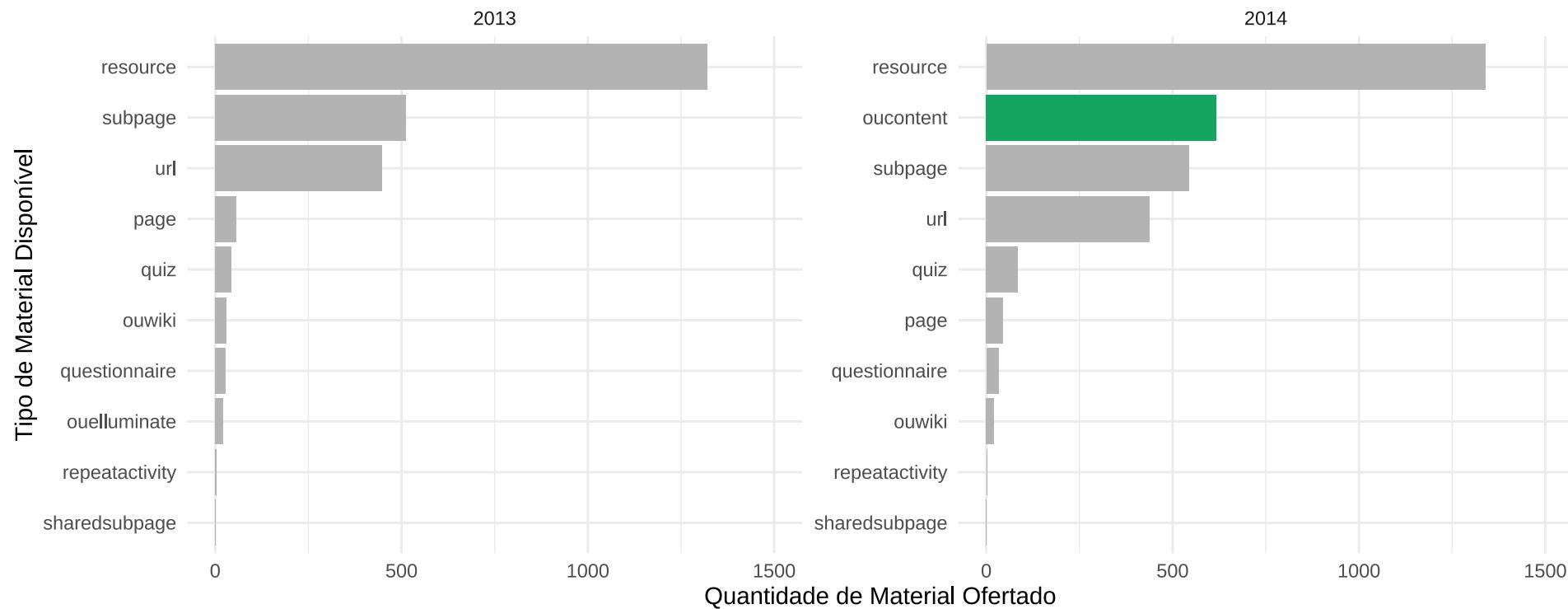
Performance vs Duração

Mesmo com uma redução considerável na carga horária dos cursos (**30 a 40 dias**), no geral a performance dos estudantes não sofreu nenhuma mudança significativa. Porém, observamos uma queda de mais de 12% quando olhamos para o **Curso B**, no período de 2014/02.



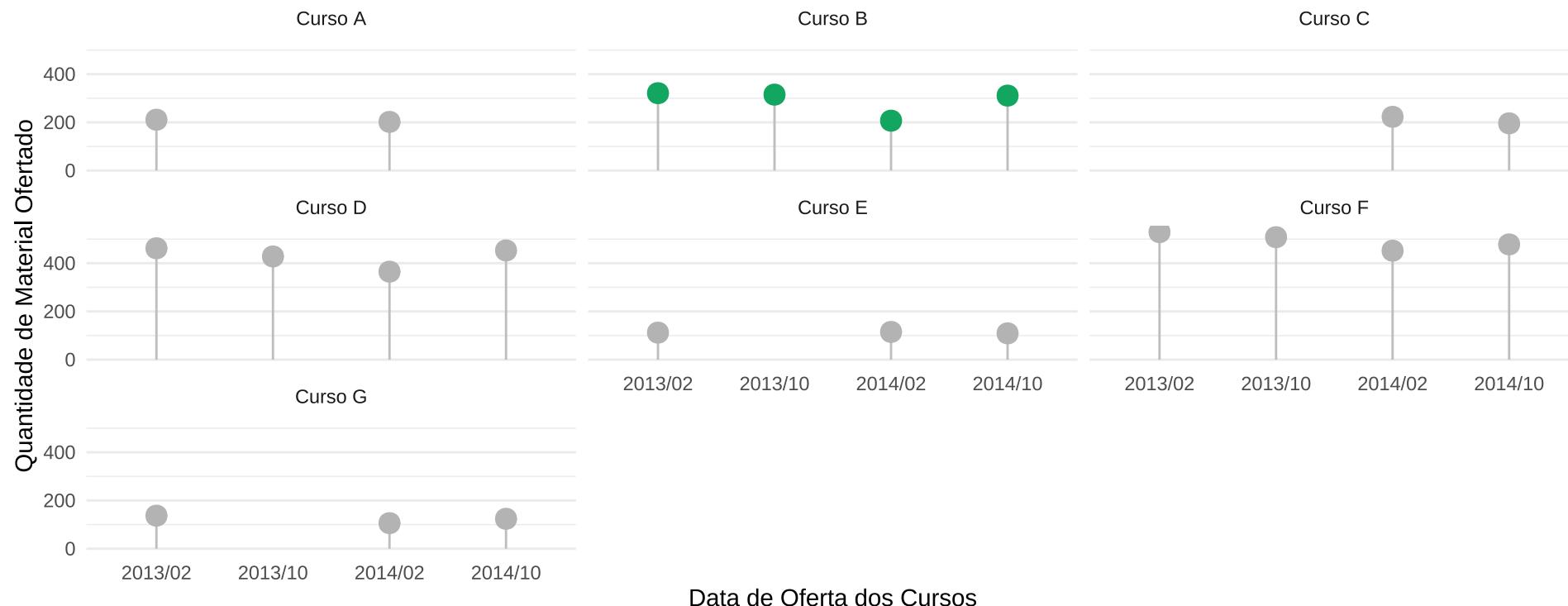
Material de consulta

A OU (Open University) oferece diversos materiais para auxiliar os estudantes. Com o tempo, houve aumento da quantidade de recursos assim como a **adição de novos**.



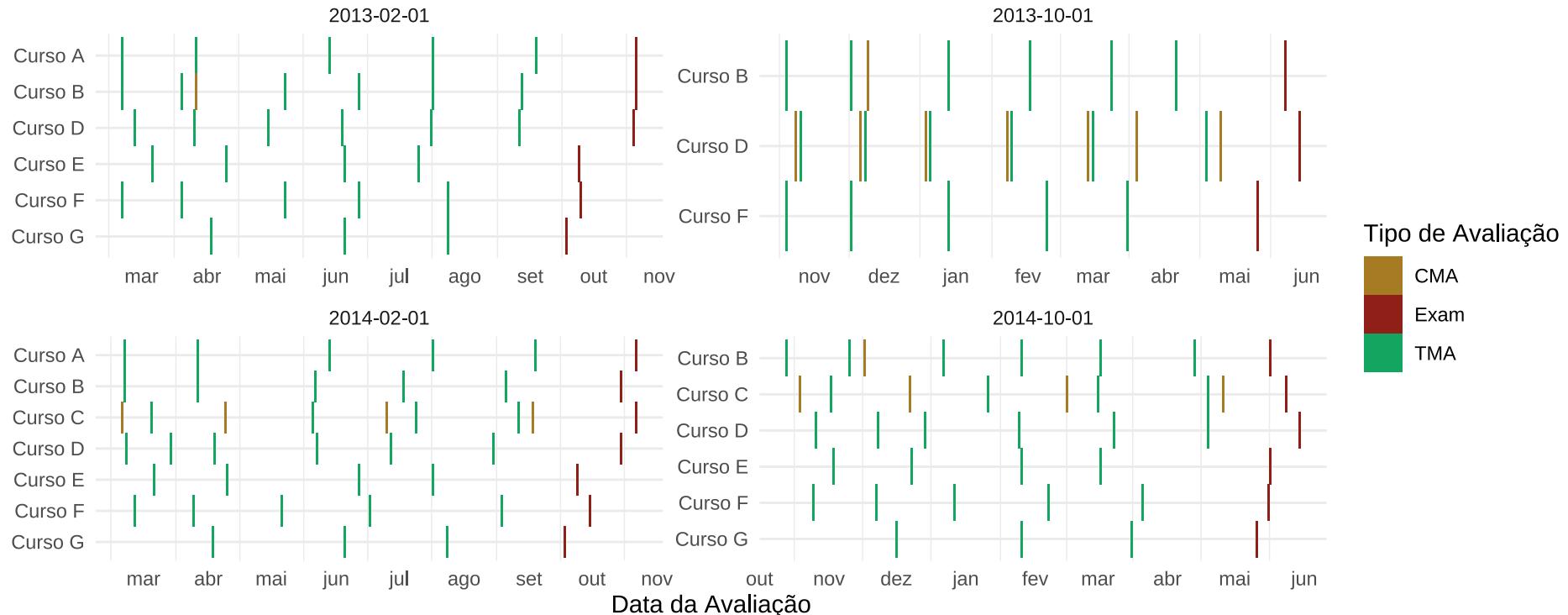
Material de consulta

Apesar do aumento na quantidade total de materiais ofertados ao longo do tempo: **2463** em 2013 e **3124** em 2014, esse comportamento não é linear quando analisamos curso a curso. E curiosamente, no curso B, no mesmo período que houve menos oferta de material (**mais de 100 recursos a menos**), é o período que observamos uma queda na média do score final dos estudantes.



Período das Avaliações

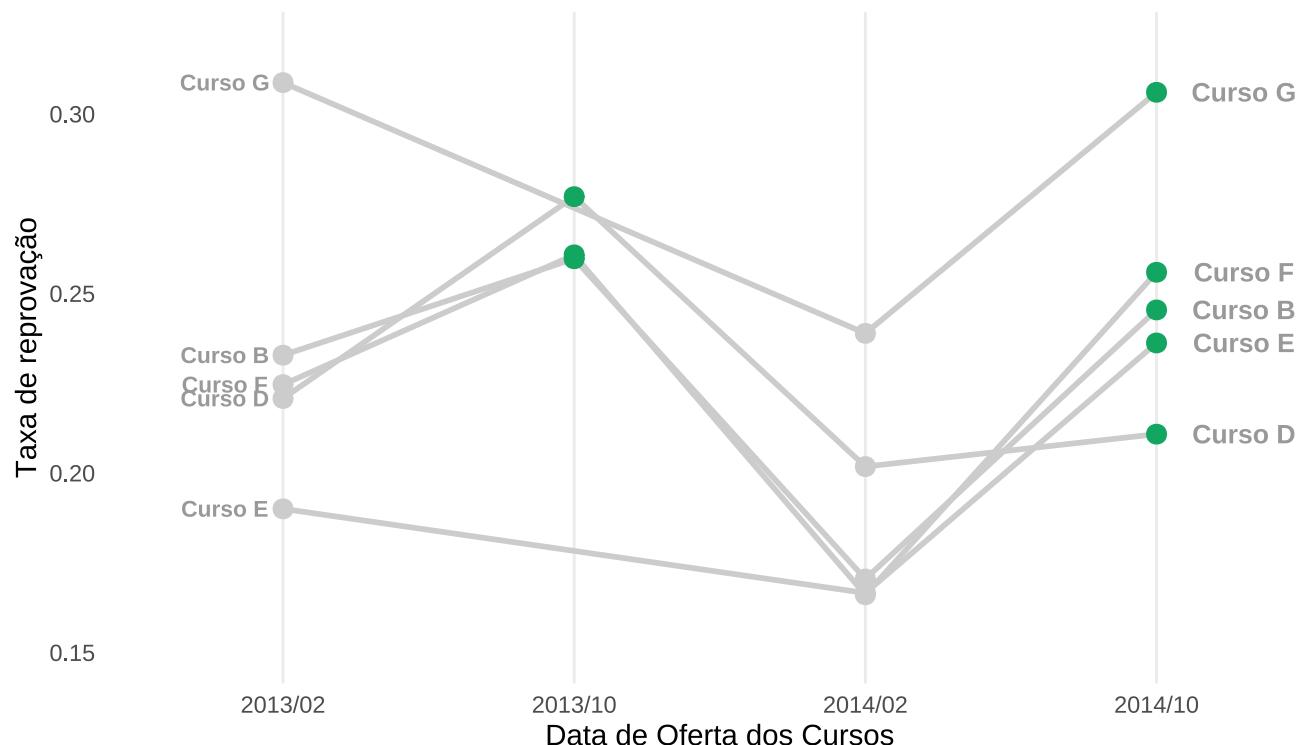
Cada curso possui avaliações intermediárias (**TMA**) ao longo do período e avaliações finais (**Exam**) na última semana. Porém, um tipo de avaliação (**CMA**) não ocorre para todas as disciplinas e nem em todos períodos. Por ser uma avaliação assistida por computador, presume-se que não se aplica a todos os cursos.



Estudantes

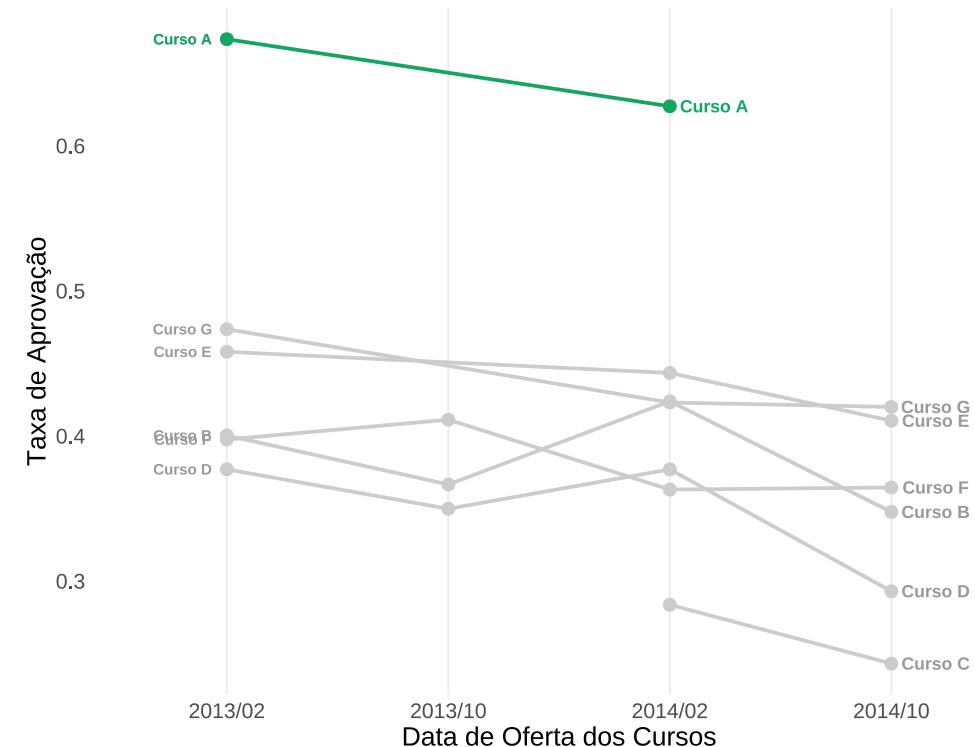
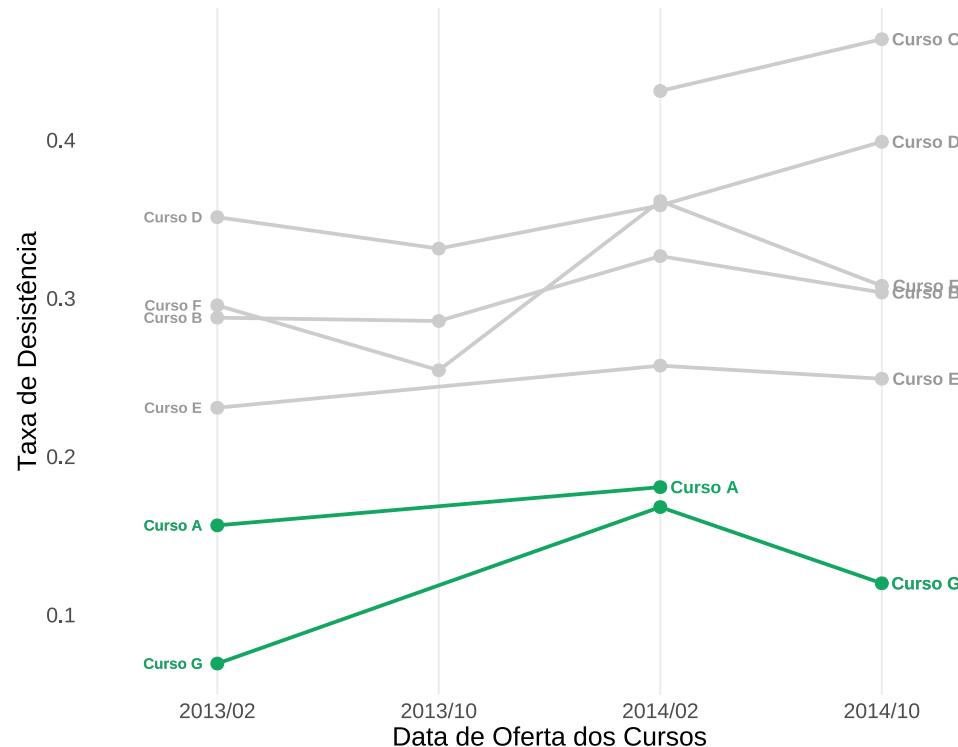
Taxa de Reprovação

Aqui observa-se um padrão interessante em relação a taxa de reprovação nos cursos. Apesar de vermos que a média de performance permanece praticamente sem alteração ao longo dos períodos (slide 5), a **taxa de reprovação** costuma ser um pouco **maior** nos períodos mais curtos (considerando os cursos que foram ofertados mais de 3 vezes).



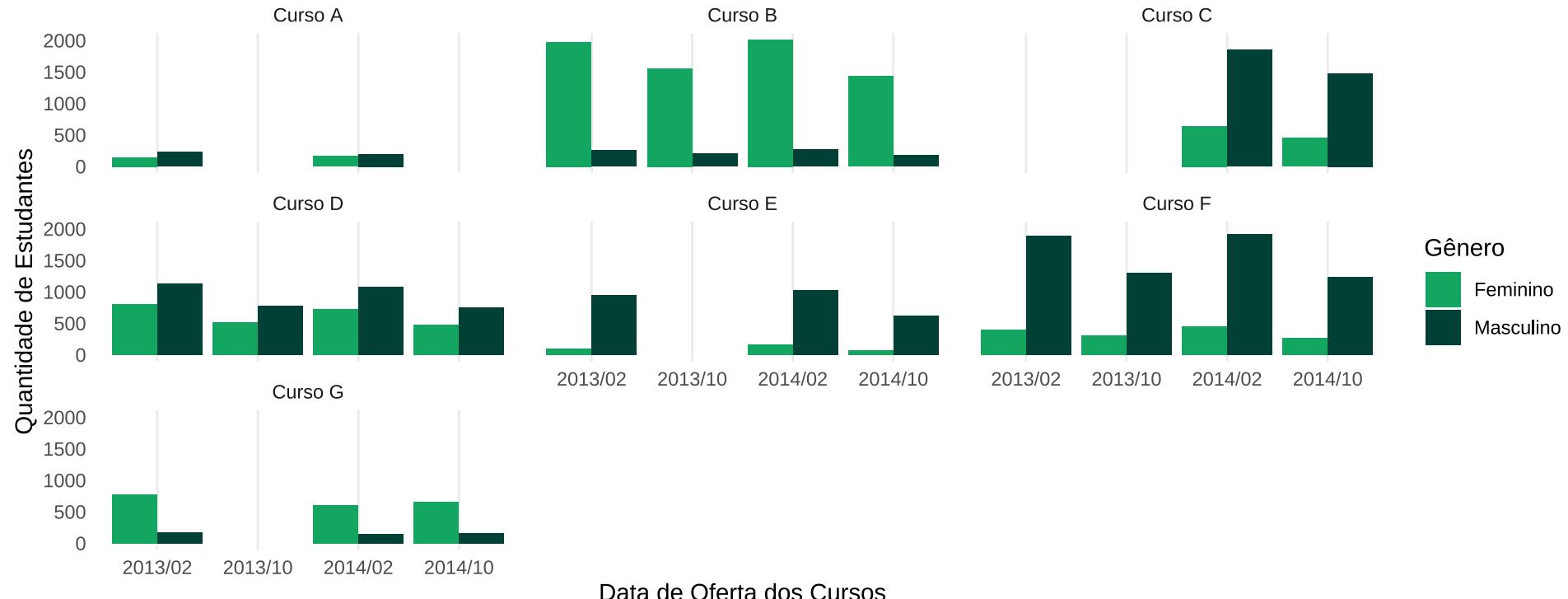
Taxas de Desistência e Aprovação

Analizando as taxas de desistência e aprovação é observado que os cursos **A** e **G** apresentam as menores taxas de desistência. Além disso, o curso **Curso A** apresenta a maior taxa de aprovação, aproximadamente 7 a cada 10 estudantes são aprovados.



Quantidade de Estudantes por Gênero

Os dados mostram uma população predominantemente feminina em 2 cursos específicos: **Curso B e Curso G**. Em todas os outros temos mais homens participando dos cursos. Tal aderência pode ter relação com características intrínsecas dos respectivos cursos.



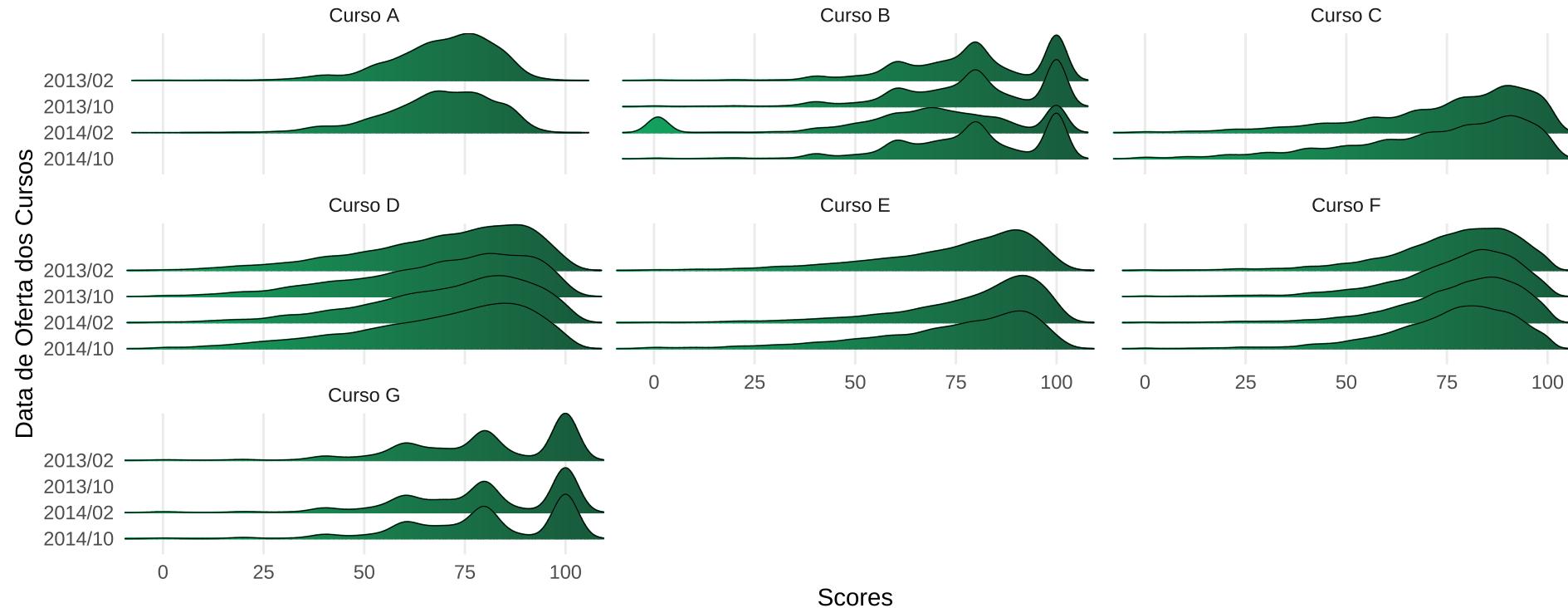
Distribuições dos Scores Finais

Após ter visualizado no slide 5 a média dos scores finais, é interessante observar como foi a distribuição de notas ao longo do tempo e por curso.

O que observa-se é um comportamento diferenciado em 2 cursos (**Curso B e Curso G**). Ambos possuem um pico de estudantes com performance muito elevada (**acima de 90 pontos**).

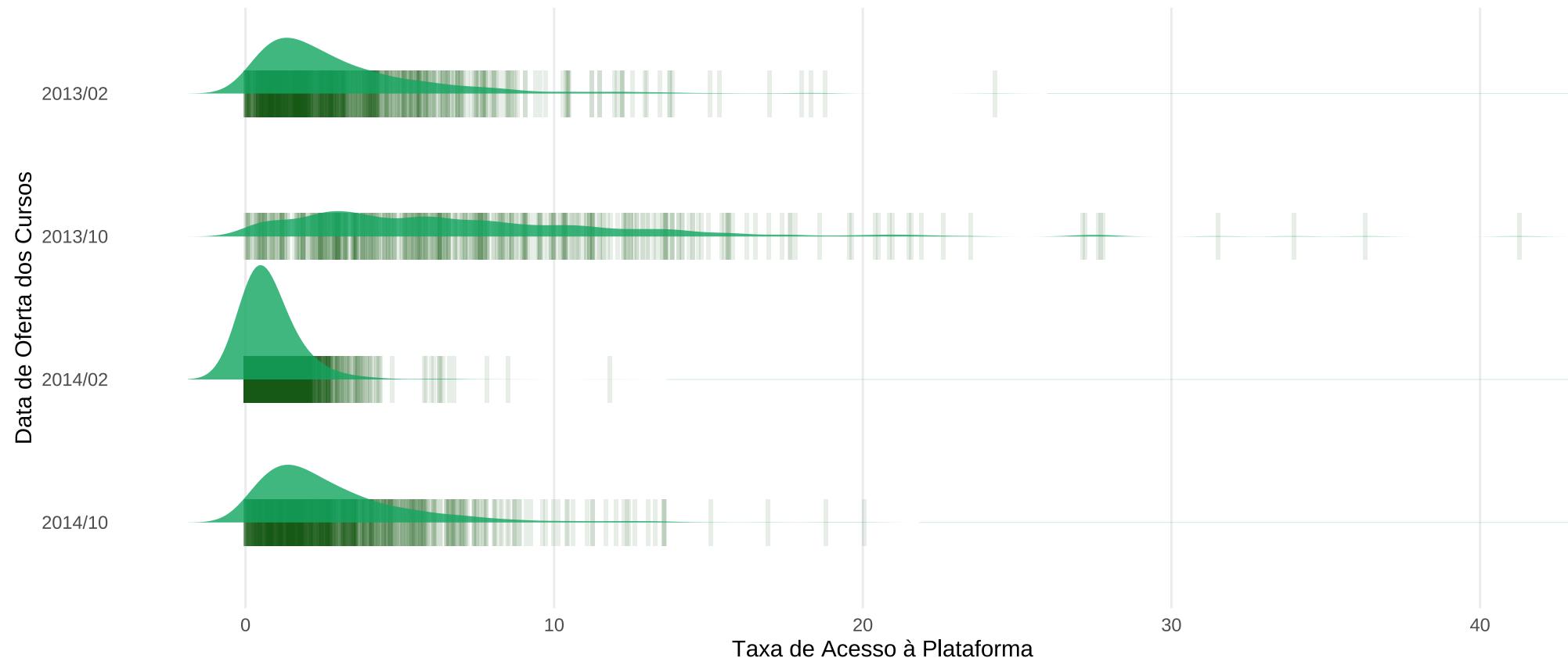
Distribuições dos Scores Finais

Também é possível notar que todos os cursos mantém um perfil característico de distribuição dos scores ao longo do tempo. Tal fator nos leva a acreditar que existe uma manutenção da forma como os cursos vem sendo trabalhados e também uma manutenção do perfil de aluno que normalmente opta por cada curso.



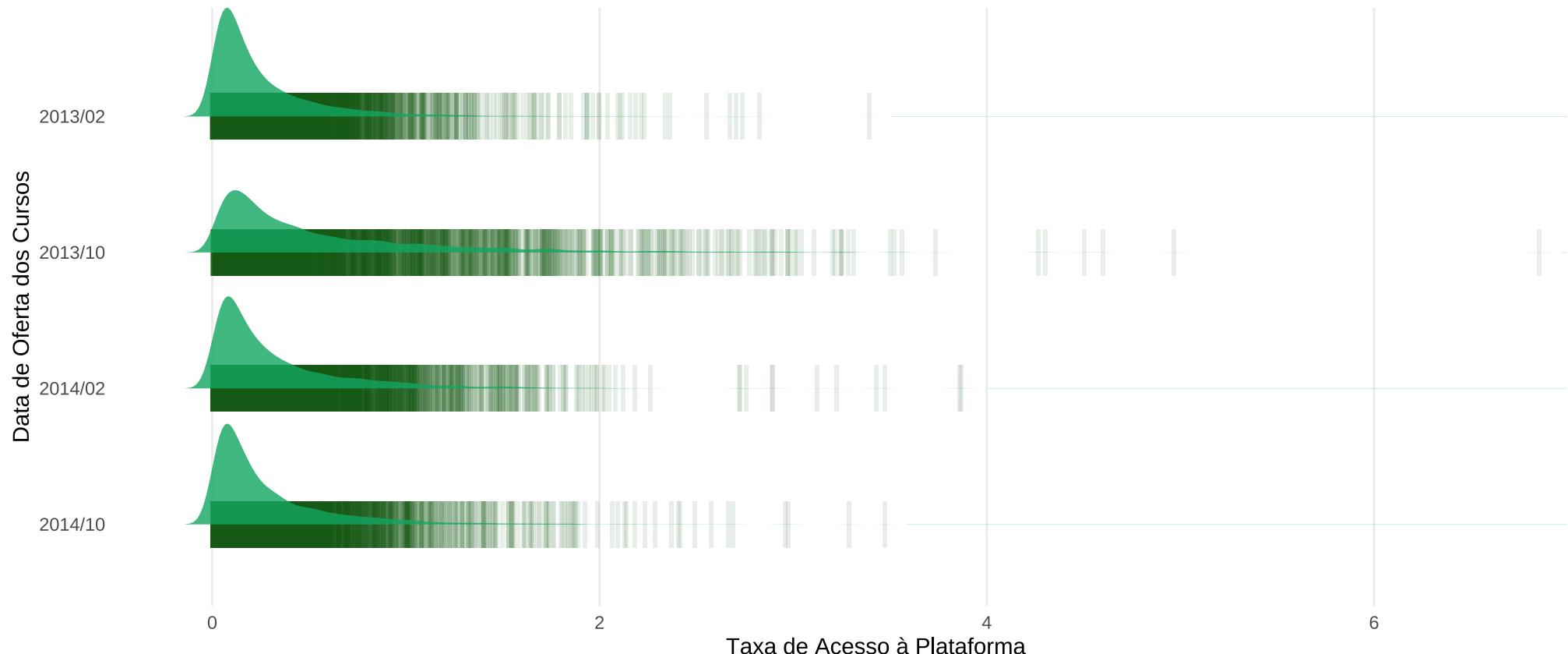
Distribuição da Taxa de Acesso à Plataforma

Como a plataforma online da OU oferece grande quantidade de materiais, e toda interação online é registrada no banco de dados, pode ser útil observar se existe diferença entre os estudantes de **alto desempenho** (High), apresentado no gráfico abaixo, e os demais.



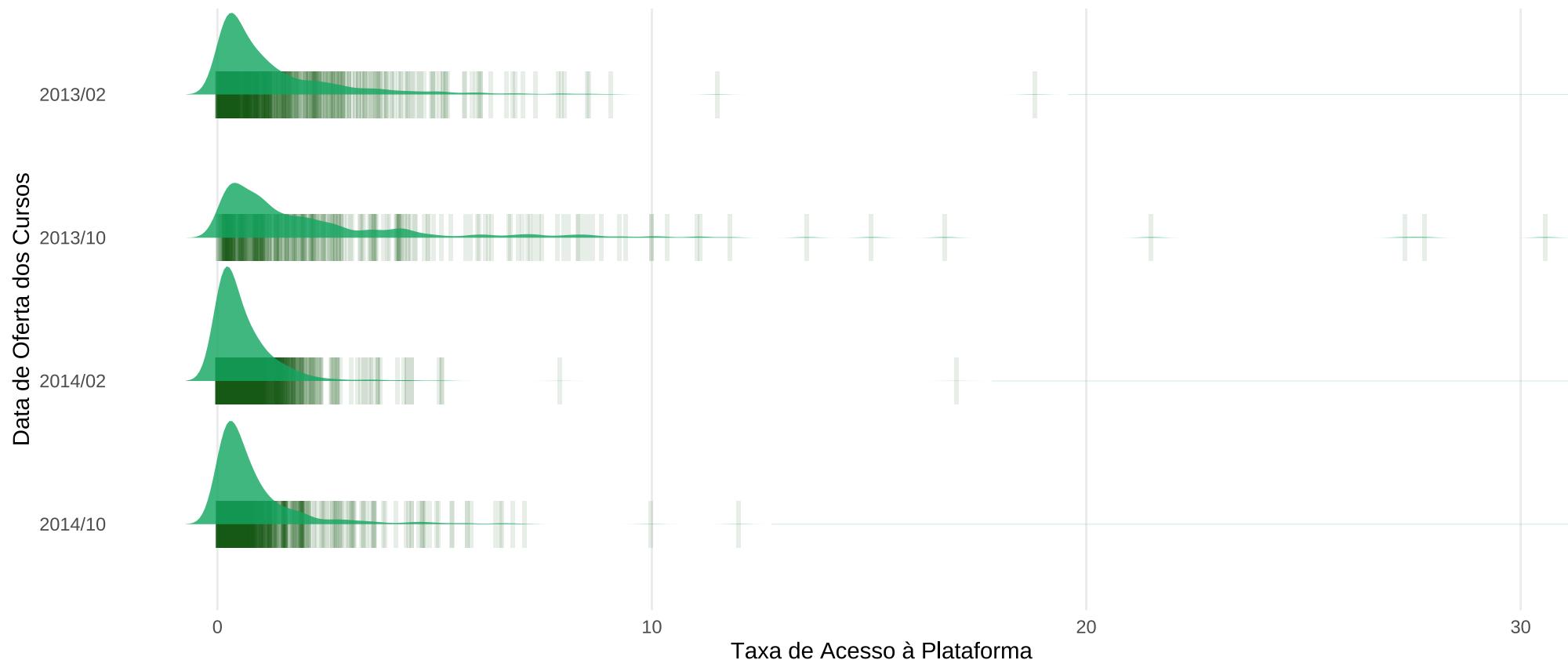
Distribuição da Taxa de Acesso à Plataforma

Abaixo estão as distribuições do acesso de estudantes que tiveram **scores entre 50 e 90** (Medium). Observando o eixo da taxa de acesso à plataforma, vemos que a mesma é muito baixa quando comparada com os estudantes de alto desempenho.



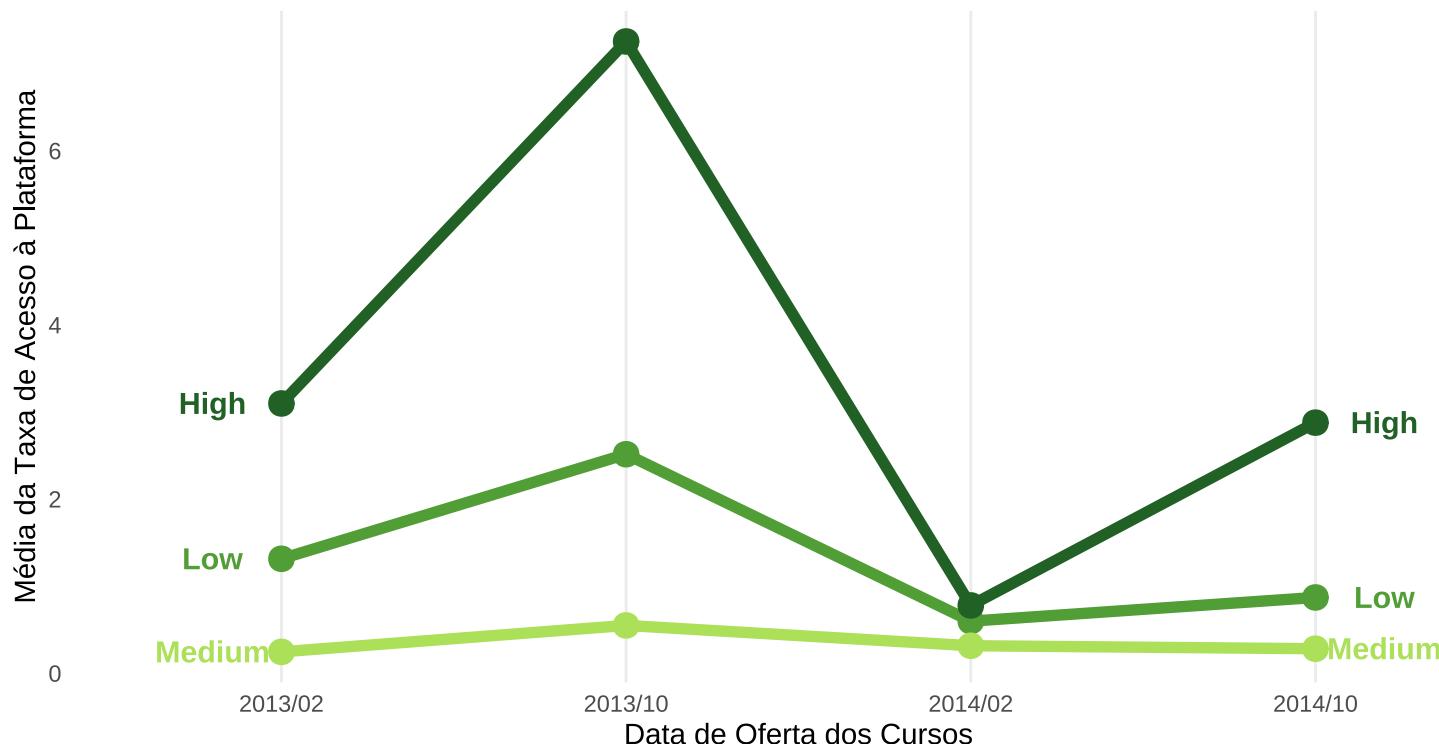
Distribuição da Taxa de Acesso à Plataforma

Abaixo estão os estudantes que foram reprovados, pois tiveram um **score menor que 50** (Low). Observa-se que mesmo sendo reprovados eles obtiveram uma taxa de acesso não tão distante dos estudantes de alto desempenho. O que leva acreditar que sua reprovação pode ser consequência de um outro fator.



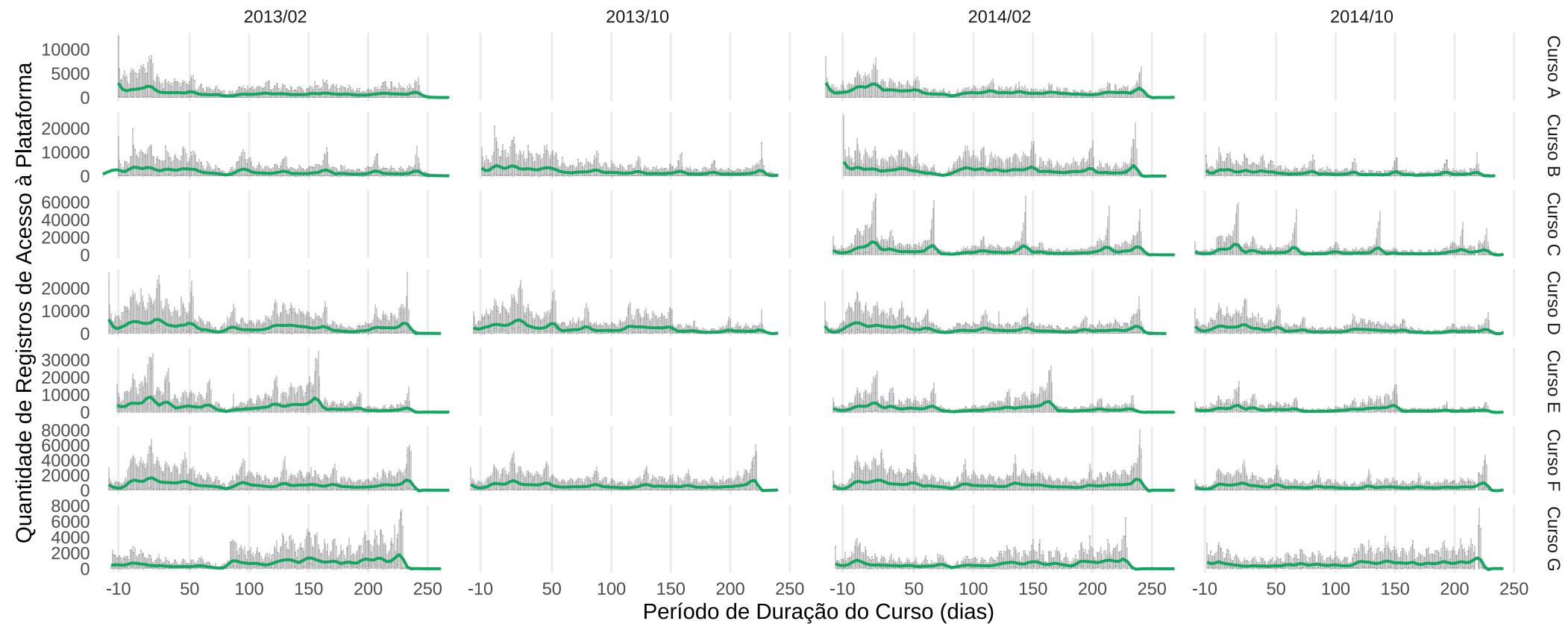
Média da Taxa de Acesso à Plataforma

Após ter uma visão geral do comportamento do acesso à plataforma entre diferentes perfis de estudantes, fica claro ao ver as médias que quem mais acessa os recursos oferecidos pela OU são os estudantes de alto desempenho. Também é possível ver que nos períodos de menor carga horária tem-se um aumento na taxa de acesso.



Perfil dos Acessos à Plataforma

No gráfico seguinte é visto o acesso à plataforma relacionado aos dados ao longo da duração do curso no eixo horizontal (em dias), e em conjunto com os cursos e o respectivo período.



Perfil dos Acessos à Plataforma

Veja que a **linha de tendência** apresenta diferentes picos durante um determinado período. Esses picos são referentes aos períodos das avaliações, sugerindo que os estudantes acessam mais os recursos quando se aproxima das datas das avaliações.

Características por Região

Observa-se insights interessantes sobre a localização desses estudantes. No **Sul do Reino Unido** está localizado a **maior parcela** dos estudantes que estiveram matriculados na OU e em segundo lugar a Escócia.

Também nota-se que existe maior número de homens do que mulheres, aproximadamente **36% a mais**.

Vemos que a maior fatia dos estudantes possuem entre **0 e 35 anos**, e novamente o Sul e a Escócia concentra maior parcela.

Auxiliando na Tomada de Decisão!

- Pode ser interessante um replanejamento dos cursos de menor duração, e assim tentar diminuir a taxa de reprovação nos respectivos períodos.
- Alguns cursos precisam ser estudadas em mais detalhes:
 - Cursos B e G: Entender quais fatores são mais determinantes para as mulheres optarem por esses cursos.
 - Curso A: Avaliar quais parâmetros fazem este curso ter altas taxas de aprovação, e assim tentar replicar para outros cursos.



Auxiliando na Tomada de Decisão!

- Os dados mostram que existe maior acesso à plataforma por estudantes de alto desempenho. Sendo assim, pode ser importante o incentivo mais efetivo ao uso da plataforma online.
- Os perfis de acesso à plataforma mostram picos de acesso ao longo de um determinado período e curso. Esses perfis podem ser importantes para o time de Data Science prever a demanda de acesso à plataforma, evitando assim quedas no servidor devido ao aumento tráfego.
- Algumas ações de marketing podem ser importantes para promover avanço nos negócios:
 - Promover o curso em outras regiões do Reino Unido, além do Sul e da Escócia, pois esses concentram maior parte dos estudantes.
 - Promover os cursos para pessoas de faixa-etária acima dos 35 anos, assim como para o público feminino.

Links

[LinkedIn](#)

[GitHub](#)

[Blog](#)

[Fonte dos dados](#)

Obrigado!