See you at the posterior line

An online racing game to teach Bayesian data analysis

Federica Zoe Ricci and Mine Dogucu (UC Irvine)

eCOTS 2024

Moving Forward with teaching Statistics and Data Science

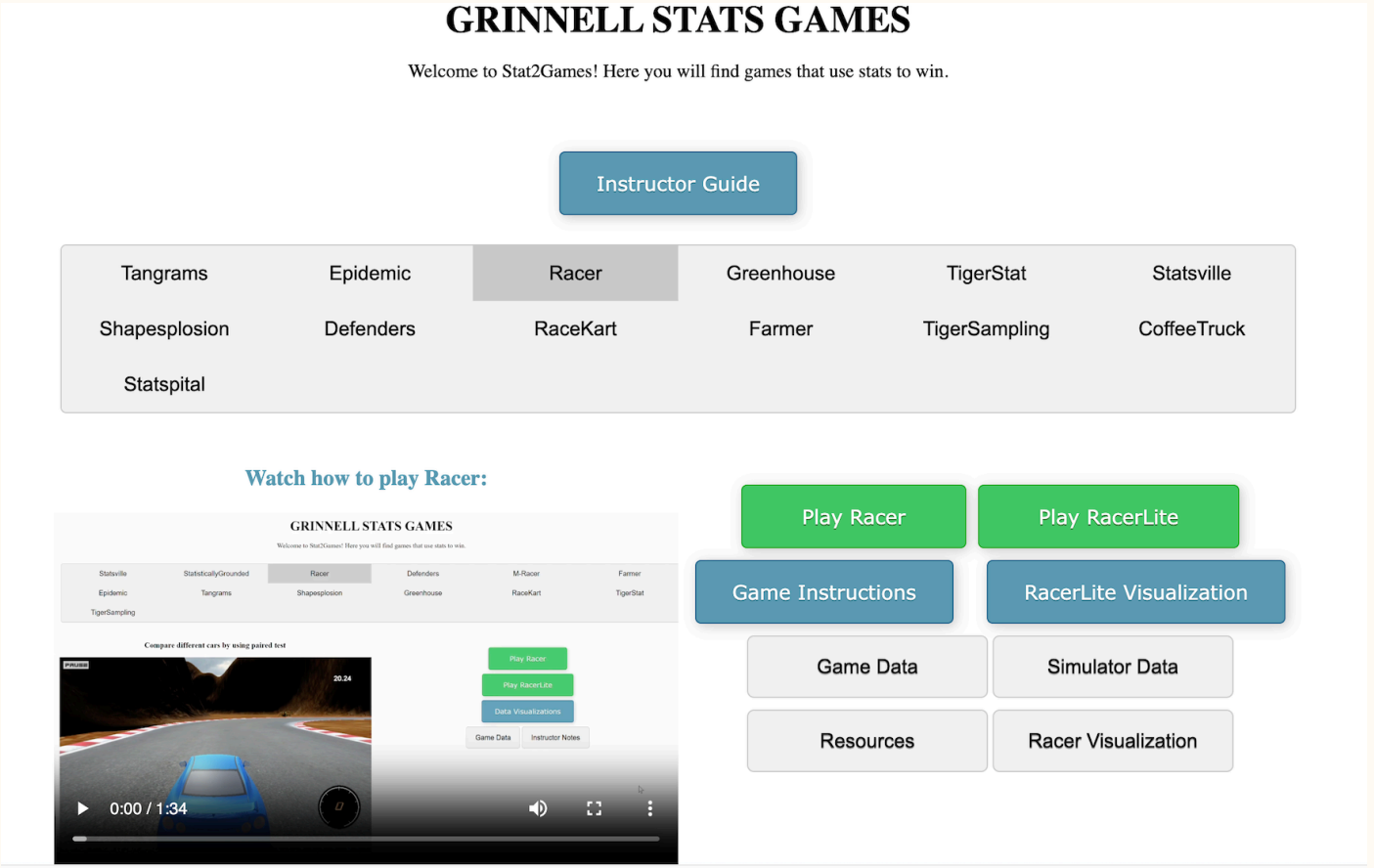
- Using data that is more **relevant** to students[1]
- More use of **fun**[2]
- More teaching **Bayes**[3]

Challenges of teaching Bayes

- **Priors** [1]
  - Where do they come from?
  - Is there *one* correct prior?
- How can people use Bayesian analysis in **real life**? [1]

We think that **games** can help with this

The Stat2Games project



Kuiper, S., & Sturdivant, R. X. (2015). Using online game-based simulations to strengthen students' understanding of practical statistical issues in real-world data analysis. The American Statistician, 69(4), 354-361.

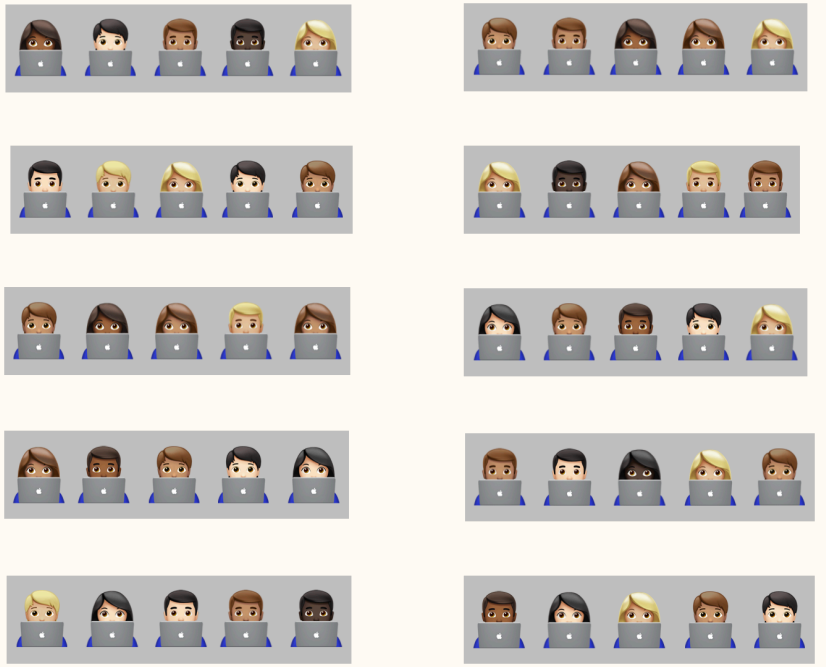
George, T., Chakraborty, A., Kuiper, S. Improving students' communication about data using online statistical games. USCOTS Workshop (2023)

Our Activity: Context

- **Introduction to Bayesian Data Analysis**
- 50 upper-level undergraduates
- 50-minute discussion (lab)
- Beta-Binomial model **already taught** in lecture

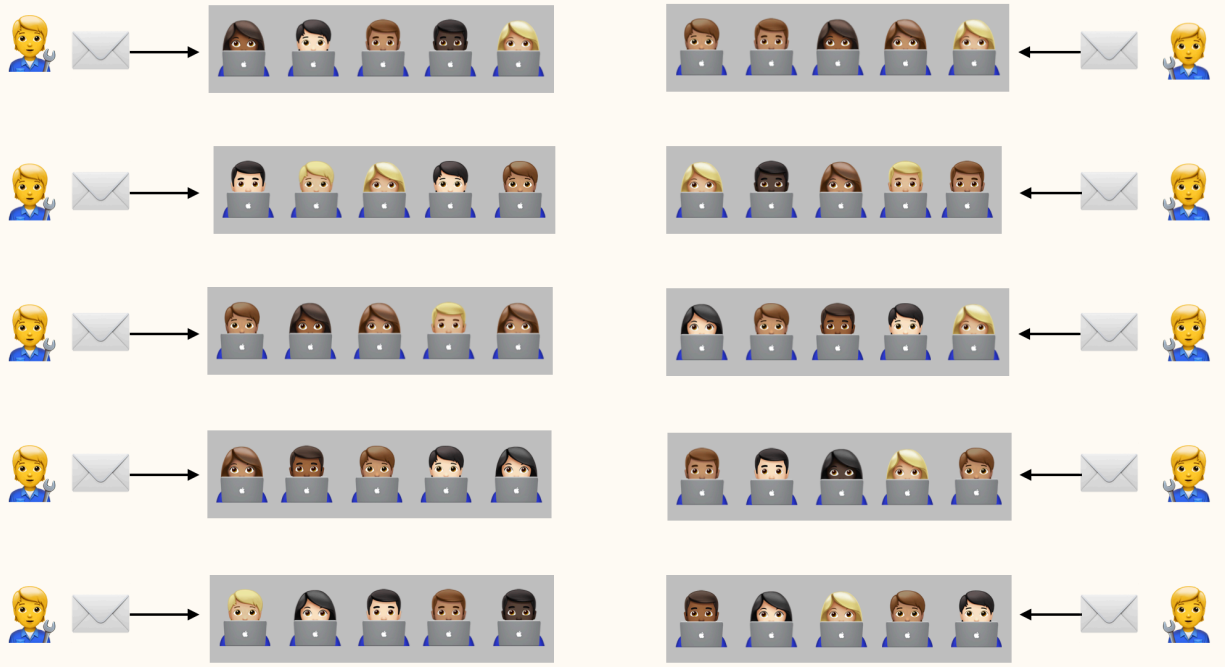
Our Activity: Setup

- Teams of 4-5 students



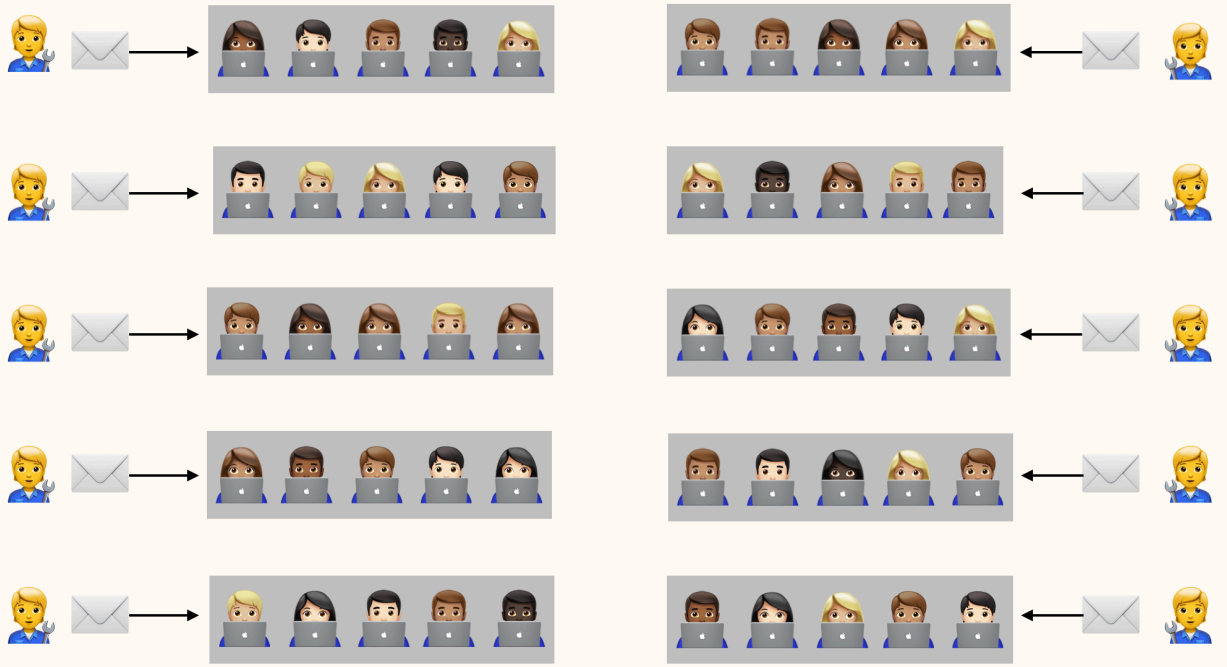
Our Activity: Setup

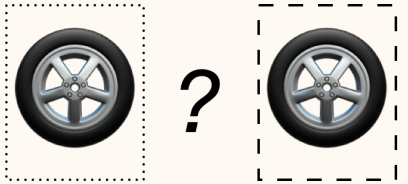
- Teams of 4-5 students
- Letters from *racing managers*



Our Activity: Setup

- Teams of 4-5 students
- Letters from *racing managers*
  - Which of two tires for an upcoming race ?

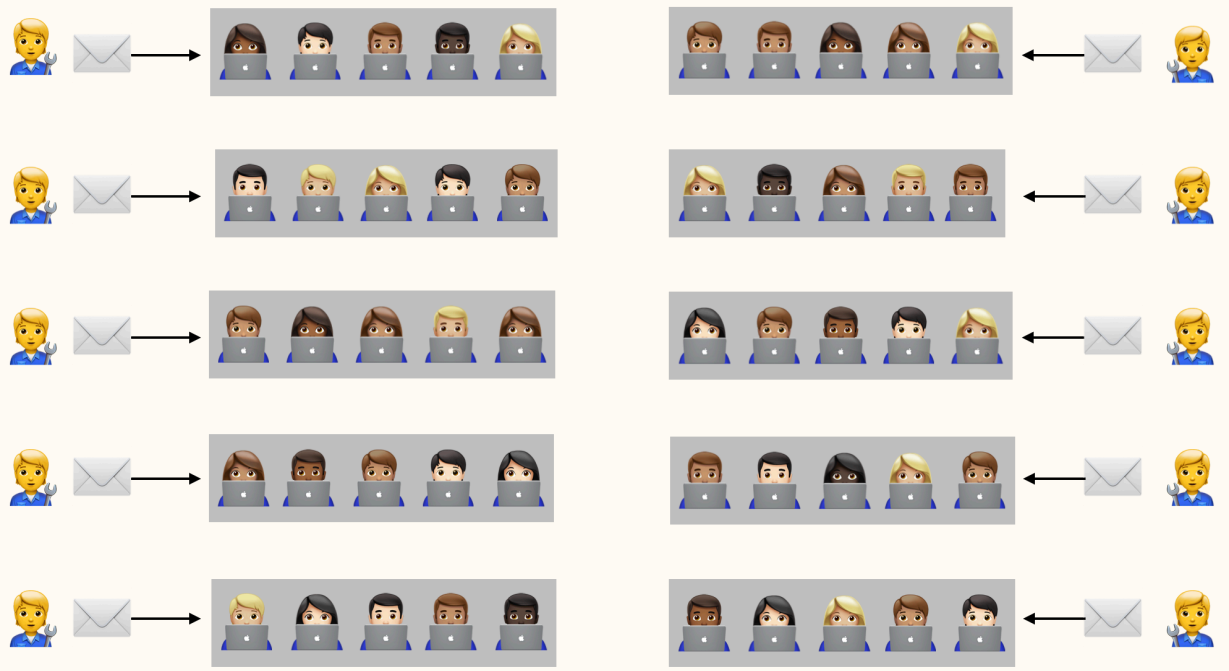


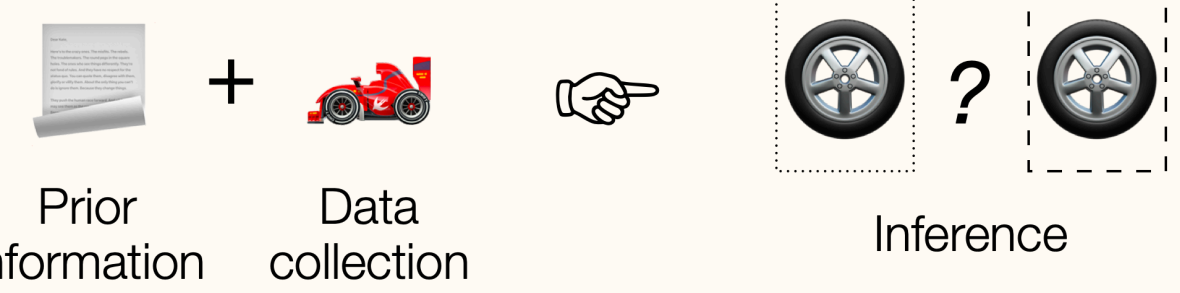


Inference

Our Activity: Setup

- Teams of 4-5 students
- Letters from *racing managers*
  - Which of two tires for an upcoming race ?
  - Combine: *engineers' believes* and *racing results* of team members





Our Activity: Prior Information

Hi team!

The race on the Eight track is approaching. This year we are going to race with the HotRod car, but we haven't finalized our choice of tires. We narrowed down the choice to the HotRod tires or the Tiny tires.

Last year, we raced with the HotRod car on a different track, the Straight track, and we were almost always faster with the Tiny tires. In fact, only 2 out of 10 times our racers finished the race faster with the HotRod tires.

Because this year the track is different, our engineers are not sure that racing with the Tiny tires would still be faster, as they might have less friction when turning. They believe it's not very likely that the HotRod tires would be faster less than 2 out of 10 times .. but that there is some chance it could actually happen up to 6 out of 10 times on the Eight track.

I would like you to collect some more data and give me some recommendations, based on what you find while playing but also based on our past experience and what our engineers believe.

Our Activity: Prior Information

Hi team!

CONTEXT

The race on the Eight track is approaching. This year we are going to race with the HotRod car, but we haven't finalized our choice of tires. We narrowed down the choice to the HotRod tires or the Tiny tires.

PRIOR DATA

Last year, we raced with the HotRod car on a different track, the Straight track, and we were almost always faster with the Tiny tires. In fact, only 2 out of 10 times our racers finished the race faster with the HotRod tires.

ENGINEERS BELIEVES

Because this year the track is different, our engineers are not sure that racing with the Tiny tires would still be faster, as they might have less friction when turning. They believe it's not very likely that the HotRod tires would be faster less than 2 out of 10 times .. but that there is some chance it could actually happen up to 6 out of 10 times on the Eight track.

DIRECTIONS

I would like you to collect some more data and give me some recommendations, based on what you find while playing but also based on our past experience and what our engineers believe.

Our Activity: Discussion Handout

Raw Quarto (template)

```
## Choosing your prior (5-10 min)

Your client asks that you take into account their prior information in your analysis.
So let's use this prior information to choose a prior distribution.
There is not a single "best" way of doing this, here we just want you to come up with something reasonable that you can explain to your client.

1.  **Briefly summarize all your prior information.**

Your answer here.

2.  **Where would you center your prior for the probability that the hot rod tires would be faster than the tiny tires? Explain your reasoning.**

Your answer here.

3.  **What are some strengths and pitfalls of your prior information?**

Your answer here.

4.  **Play around with values for** $alpha$ **and** $beta$ **below that give you a prior that looks reasonable enough to you. Briefly explain why it is reasonable (just the intuition)**.

```{r}

alpha <- 1
beta <- 1
plot_beta(alpha, beta)

```

Your answer here.

Data collection (5-10 min)

Whenever playing set **Player ID = RacingStats** and **Group ID = TeamRole**, e.g. **Group ID = Danica1Facilitator**.

First, each team member will do one trial with the Classic car on the Eight track:

<https://www.stat2games.sites.grinnell.edu/games/raceradvanced22.html>

Then each team member will race the **HotRod car** on the **Eight** track, once with the HotRod tires and once with the Tiny tires:

-   Facilitator and Annotator race *first* with the HotRod tires

-   Summarizer, Timer and Speaker (if present) race *first* with the Tiny tires

IMPORTANT: **Remember to note your finishing times!!!**

Then complete the code below.

```{r}

# Assign value 1 if the finishing time with the HotRod tires was shorter than
# your finishing time with the Tiny tires, 0 otherwise

facilitator <- 0
annotator <- 0
summarizer <- 0
timer <- 0
speaker <- 0

# Uncomment the following line if your team has 4 people
n_hotrod_faster <- facilitator + annotator + summarizer + timer
```

Rendered by one of the teams

Choosing your prior (5-10 min)

Your client asks that you take into account their prior information in your analysis. So let's use this prior information to choose a prior distribution. There is not a single "best" way of doing this, here we just want you to come up with something reasonable that you can explain to your client.

1. **Briefly summarize all your prior information.**

Its a beta distribution, last year the HotRod tires raced faster than the tiny tires 2 out of 10 times. The tiny tires are faster on a straight track. This year, with a different track, we expect the HotRod tires to race faster up to 6 out of 10 times.

2. **Where would you center your prior for the probability that the hot rod tires would be faster than the tiny tires? Explain your reasoning.**

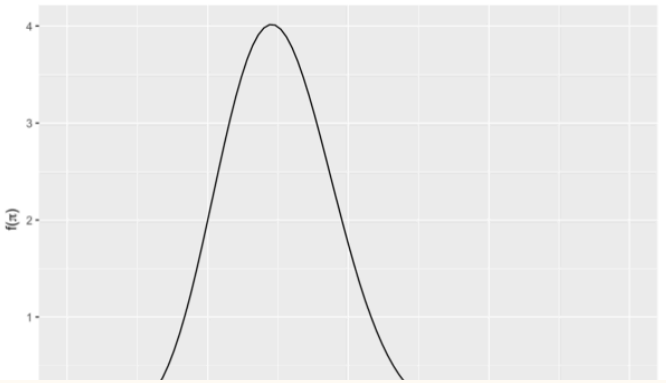
We would center the prior around 0.4 because it is between last year's performance and this year's peak expectation.

3. **What are some strengths and pitfalls of your prior information?**

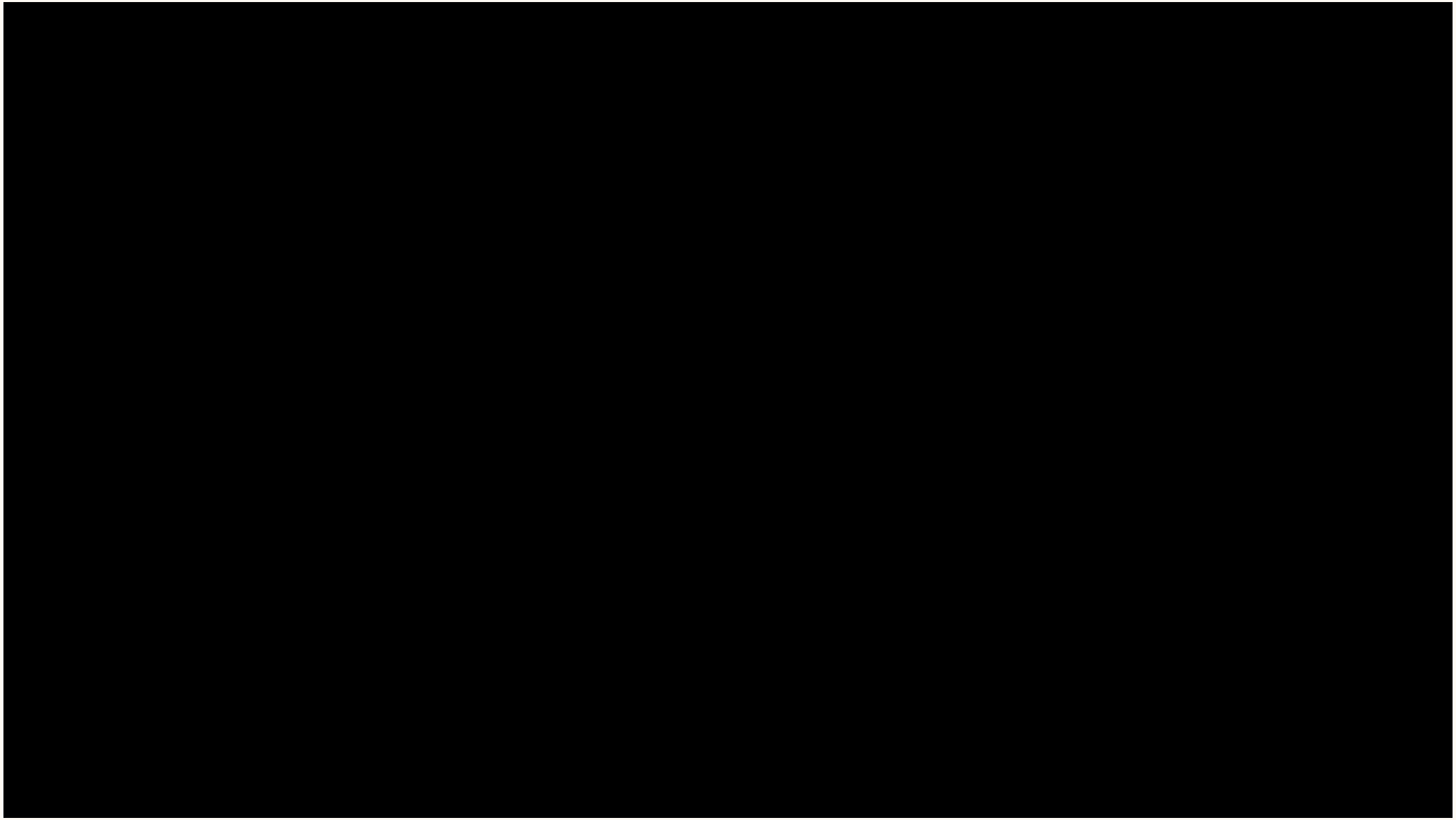
A strength is that we understand part of why the HotRod tires were slower from experience, a pitfall is that we have no data about the HotRod tires on a different track.

4. **Play around with values for $alpha$ and $beta$ below that give you a prior that looks reasonable enough to you. Briefly explain why it is reasonable (just the intuition)**.

```
alpha <- 9
beta <- 15
plot_beta(alpha, beta)
```


```

Our Activity: Data Collection



Our Activity: Discussion Handout

Quarto file rendered by one of the teams

Choosing your prior (5-10 min)

Your client asks that you take into account their prior information in your analysis. So let's use this prior information to choose a prior distribution. There is not a single "best" way of doing this, here we just want you to come up with something reasonable that you can explain to your client.

1. **Briefly summarize all your prior information.**

Its a beta distribution, last year the HotRod tires raced faster than the tiny tires 2 out of 10 times. The tiny tires are faster on a straight track. This year, with a different track, we expect the HotRod tires to race faster up to 6 out of 10 times.

2. **Where would you center your prior for the probability that the hot rod tires would be faster than the tiny tires? Explain your reasoning.**

We would center the prior around 0.4 because it is between last year's performance and this year's peak expectation.

3. **What are some strengths and pitfalls of your prior information?**

A strength is that we understand part of why the HotRod tires were slower from experience, a pitfall is that we have no data about the HotRod tires on a different track.

4. **Play around with values for $alpha$ and $beta$ below that give you a prior that looks reasonable enough to you. Briefly explain why it is reasonable (just the intuition).**

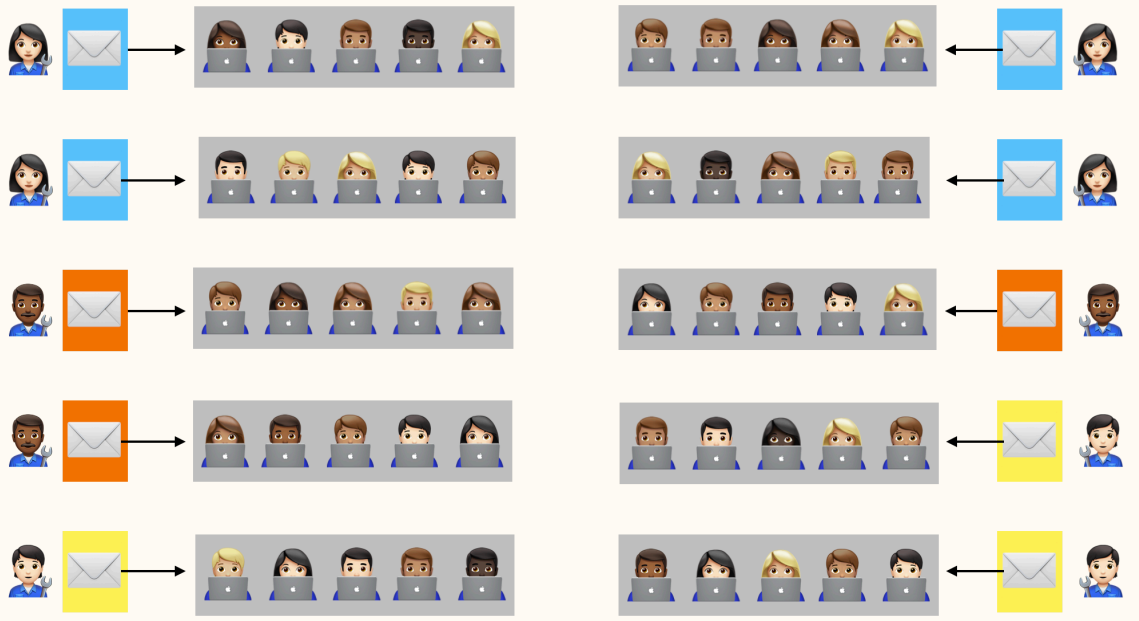
Our Activity: Learning Goals

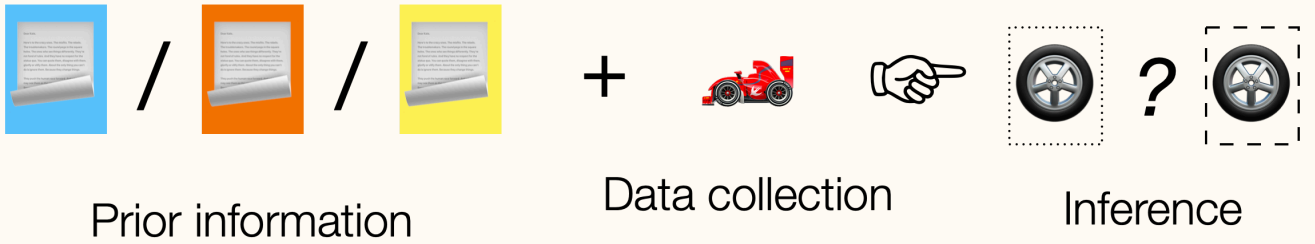
- practice using given information to **formulate a prior distribution**
- apply the Beta-Binomial model to a **fun and real-world** application

- observe that different priors for the same problem can be reasonable?

Our Activity: Setup

- Teams receive *different information* based on their managers:
  - Team Danica
  - Team Lewis
  - Team Mario
- Naturally motivates different priors





Our Activity: Prior Information

Team Danica

Hi team!

The race on the Eight track is approaching. This year we are going to race with the HotRod car, but we haven't finalized our choice of tires. We narrowed down the choice to the HotRod tires or the Tiny tires.

Last year, we had four players racing on the Eight track with the Classic car, once with the HotRod and once with the Tiny tires. Three of them finished faster with the HotRod tires, one of them with the Tiny tires. For some players this might have been their first race, and everyone raced with HotRod tires first.. That's unfortunate from your perspective, right? But let's pretend that data is okay.

By the way, our engineers think it's likely that 6 out of 10 races the HotRod tires would be faster than the Tiny ones..but they think that it could also be as low as 4 out of 10 races, or as high as 9 out of 10 races.

I would like you to collect some more data and give me some recommendations, based on what you find while playing but also based on our past experience and what our engineers believe..

Team Lewis

Hi team!

The race on the Eight track is approaching. This year we are going to race with the HotRod car, but we haven't finalized our choice of tires. We narrowed down the choice to the HotRod tires or the Tiny tires.

Last year, we raced with the HotRod car on a different track, the Straight track, and we were almost always faster with the Tiny tires. In fact, only 2 out of 10 times our racers finished the race faster with the HotRod tires.

Because this year the track is different, our engineers are not sure that racing with the Tiny tires would still be faster, as they might have less friction when turning. They believe it's not very likely that the HotRod tires would be faster less than 2 out of 10 times .. but that there is some chance it could actually happen up to 6 out of 10 times on the Eight track.

I would like you to collect some more data and give me some recommendations, based on what you find while playing but also based on our past experience and what our engineers believe.

Team Mario

Hi team!

The race on the Eight track is approaching. This is the first year we are racing on this track..I am always looking for new challenges! All players are racing with the HotRod car this year, but we can choose which tires to use.

Our engineers believe the best are the HotRod tires or the Tiny tires. But they have no strong opinion which of the two would be faster. They think it might be just as likely that the HotRod tires are faster 2 out of 10 times.. or maybe 8 out of 10 times. They exclude that one of these two tires would always be faster.

I would like you to collect some data and give me some recommendations, based on what you find while playing but also based on our engineers' opinions.

Students' perspectives

Remarks from 32 participants (open-ended questions)

- 22 students: **Fun**
- 17 students: **Stats understanding**
- 15 students: **Interactive**
- 14 students: **Teamwork**
- 9 students: **Gamification**
- 9 students: **Data collection**

> *The racing game was very fun which make learning the concepts much more engaging than simply reading about a simulation in a book.*

> *Being able to use R and see how our race results impacted the posterior model had a positive impact on my learning.*

> *The activity was fun and helped me understand the process of making models.*

> *Helped us interact with other members of the class. Car game was fun.*

Lessons we learned

🎤 18 students: **Time**

- **50 minutes not enough** to complete activity for all teams
- Definitely not enough for ~~**class discussion**~~

🎤 6 students: **Seating arrangement** in a row

- can affect **team discussion**

👍 **Overall very promising**

- This activity **really engaged** students

  > *I really liked it, if every discussion was like this I would have no problem going at 9:00am.*

Material shared

All materials developed for the activity are available at github.com/federicazoe/bayes-games:

- slides to guide the activity
- discussion handout (.qmd)
- racing team letters & more

See you at the posterior line

Live eCOTS session

[Tuesday, June 11th, 2:40 pm – 3:25 pm ET]