NFL Injuries

- 5712 unique games
- 105 injuries

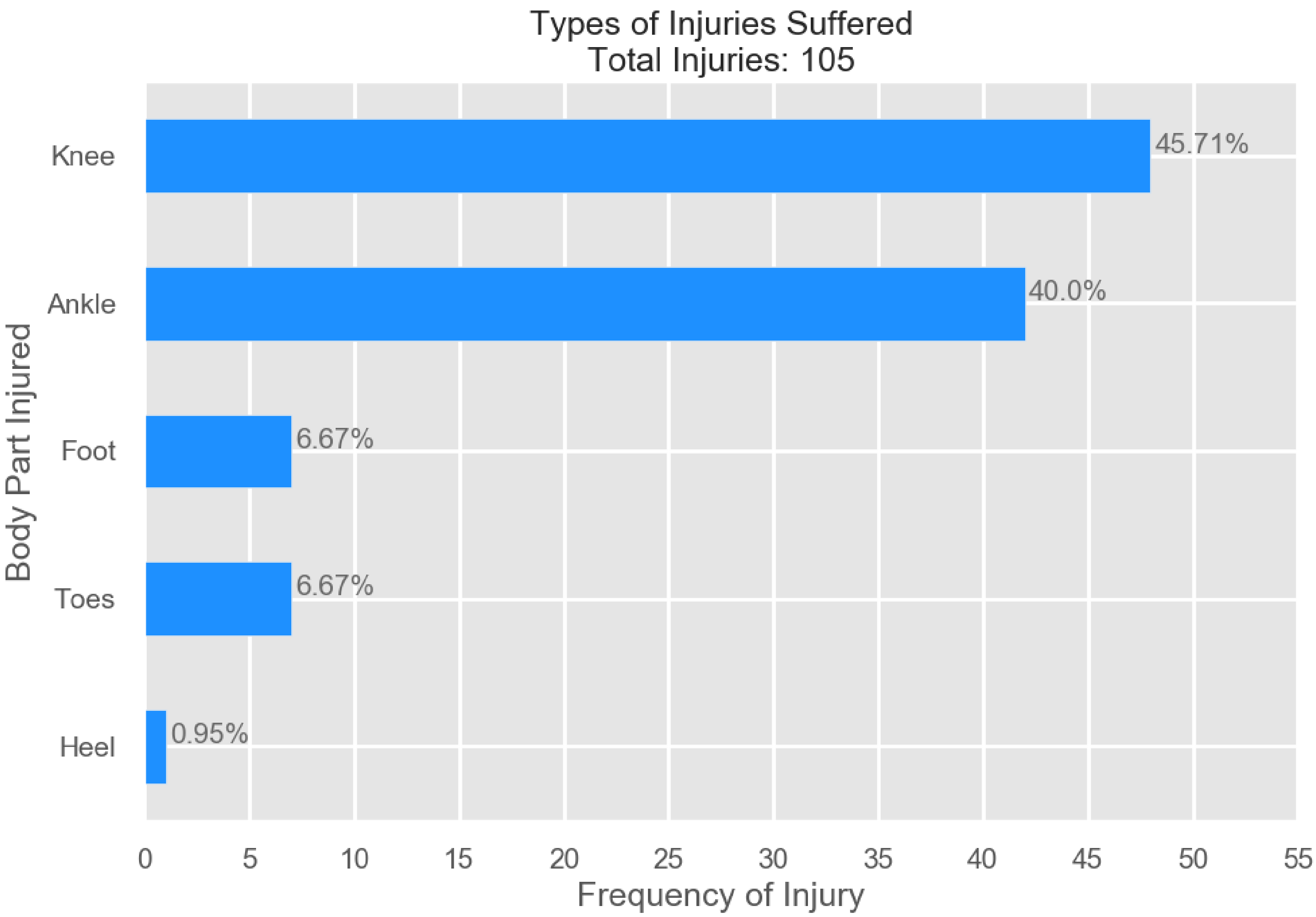
Initial ideas

- Is there a link between surface type and:
  - Length of injury
  - Body part injured
- Is there a link between body part injured and length of injury?
  - Would assume knee is the worst (and most common apparently)
- Is there a link between stadium type and field type?
  - Are more open fields synthetic due to upkeep costs?
  - If the stadium is open, does weather make injuries more common / worse
- Is there a link between position and type of injury and how long they are injured
  - Perhaps assume runners i.e. RBs, HBs, QBs that run (Lamar Jackson) would be more likely to get non-contact injuries
- If a player has already had an injury, are they more likely to get another injury?
  - With dislocation this is common, some injuries are 'repeat offenders'
- How does speed impact injury type?
  - Does faster = worse injury?
  - Does synthetic turf increase or decrease average speed?
- Is there a particular stadium that has the most injuries? If so, why would that be?
- Is there a particular team that suffers more injuries? If so, why?

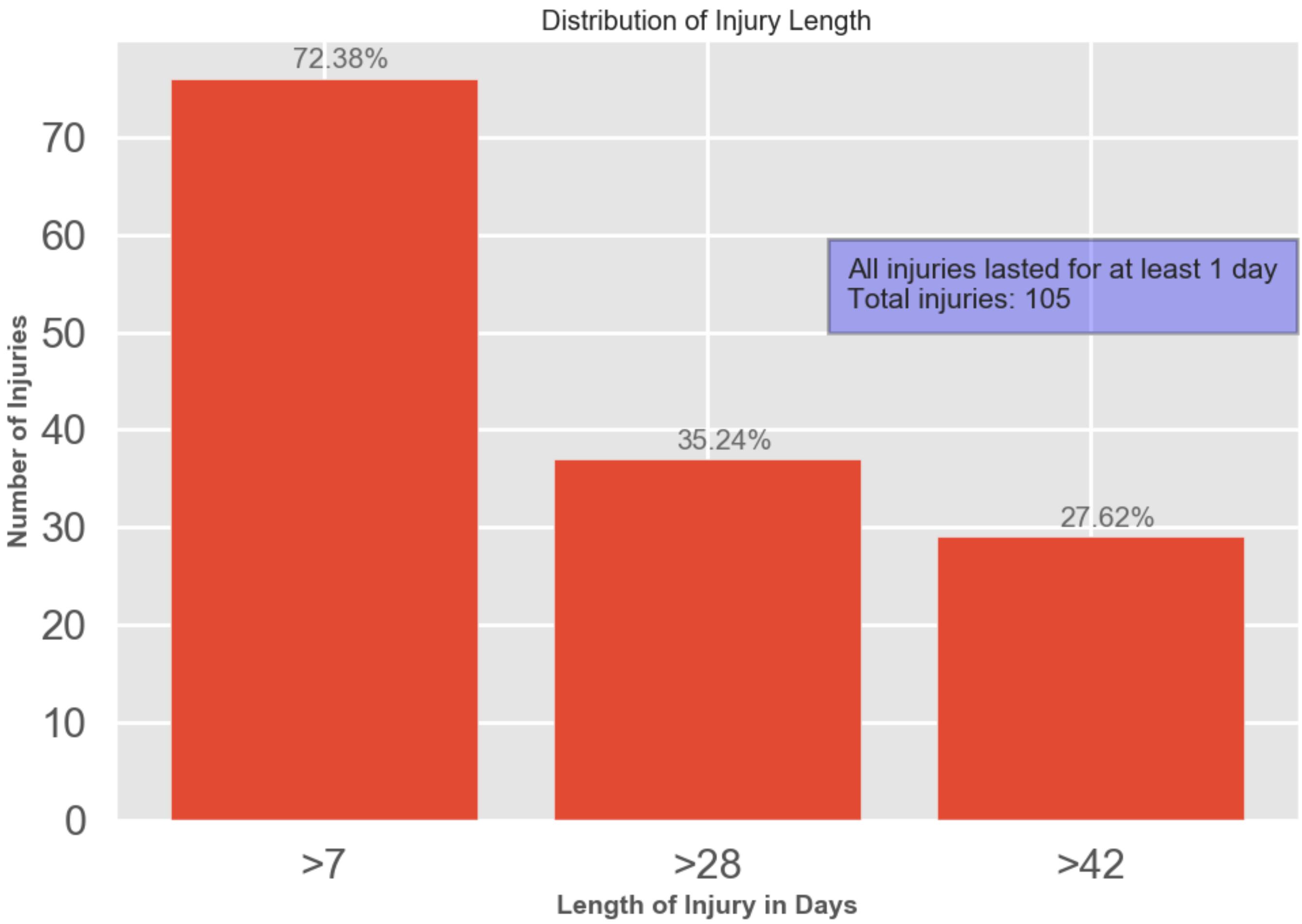
Extra

- I wonder if there is a link to the number of plays already played by the time the injury occurs
- Whats the mean, mode, median speed of players that were injured and were not
  - ^^ same for different surface types
- Is there a correllation between surface type and speed?
  - Does weather impact traction less on synthetic pitches?
    - One thing is for certain, natural pitches divet and cause fumbles and perhaps ankle injuries
    - Maybe there is an obscure link between rain + natural pitches and ankle injuries

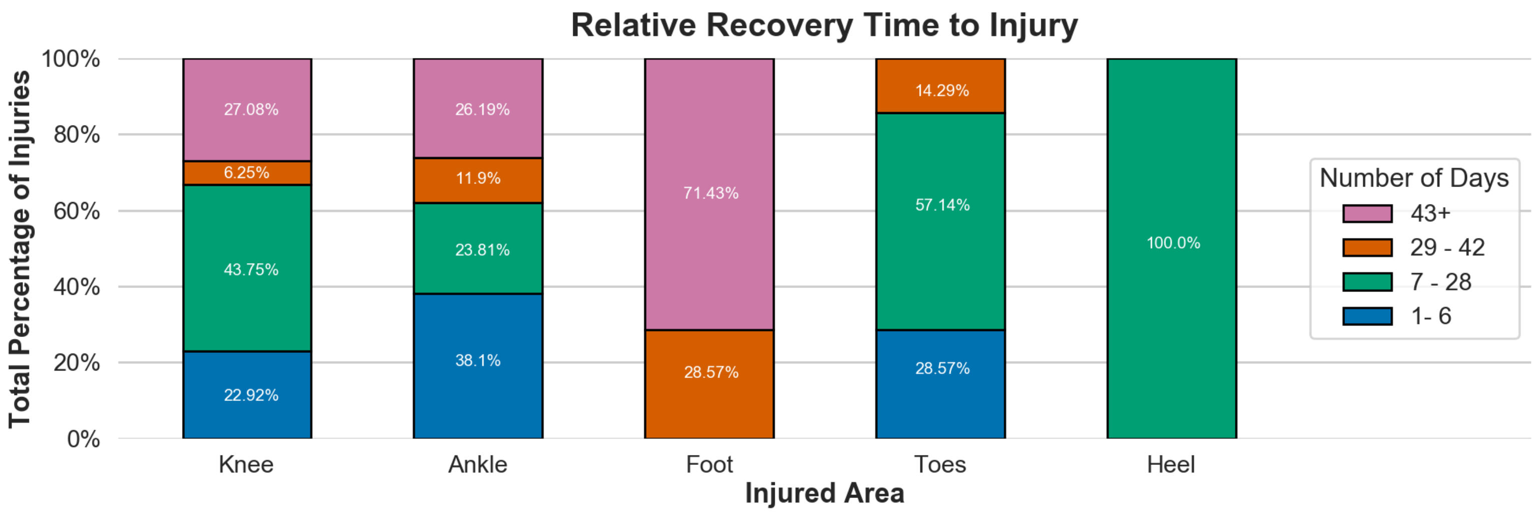
Understanding the injuries



- The most common injuries are also arguably the most severe: the ankle and the knee especially are incredibly complex joints. Some knee injuries can be career ending
- We often hear of 'torn ACL' in the NFL (Anterior Cruciate Ligament) which is a common knee injury
    - This, however, is usually a much longer-term injury and one would assume that a torn ACL MUST take longer than 42 days to recover
    - Having checked some papers, ACL rehabilitation is a very lengthy process



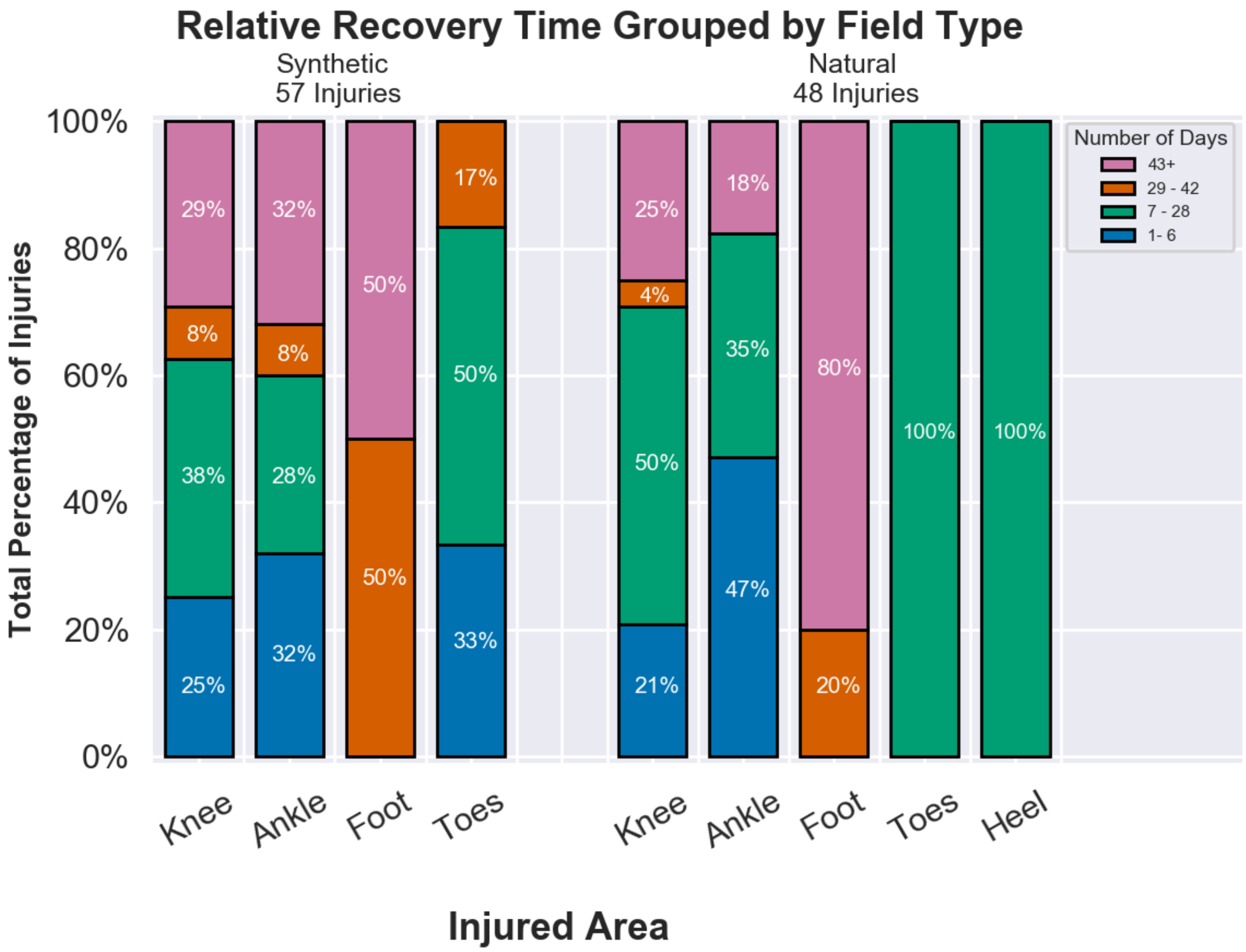
- If 72% of injuries were 7 days or longer, this also means that a number of injuries (~29) lasted between 1 day and 7 days



This graph allows for a quick estimate of recovery time per injured area and provides some interesting insight. It would suggest that Foot injuries, whilst only making up 6.67% of the total injuries, has in 70% of cases become a long-term injury (43+ days). Furthermore, any foot injury that was not in the 'long-term' band fell between 29 and 42 days which is still significant – even if we assume that a team plays Sunday games each week (no Thursday game), this would be between 4 – 6 games of the regular season. Another unexpected observation from the graph is that the Heel injury (because there is only 1) was a significant one and caused the player to miss 1 game at least. Knee and Ankle injuries arguably hold the most weight as they are by far the more prevalent type of injury in this dataset. The graph would suggest that Knee injuries tend to last longer compared to an ankle injury. Looking at this graph, I became intrigued as to the type of common injuries that may fall into each group to gain an understanding of why certain injured areas tend to take longer to recover. Before exploring the types of injuries, I aim to split the above data into the different surface types.

In one example PlayerKey: 47307 GameID: 47307-10

- Player got 2 separate injuries, one ankle, one knee



The real question is; is there a statistically significant difference between then number of injuries that occur on Synthetic pitches vs Natural pitches

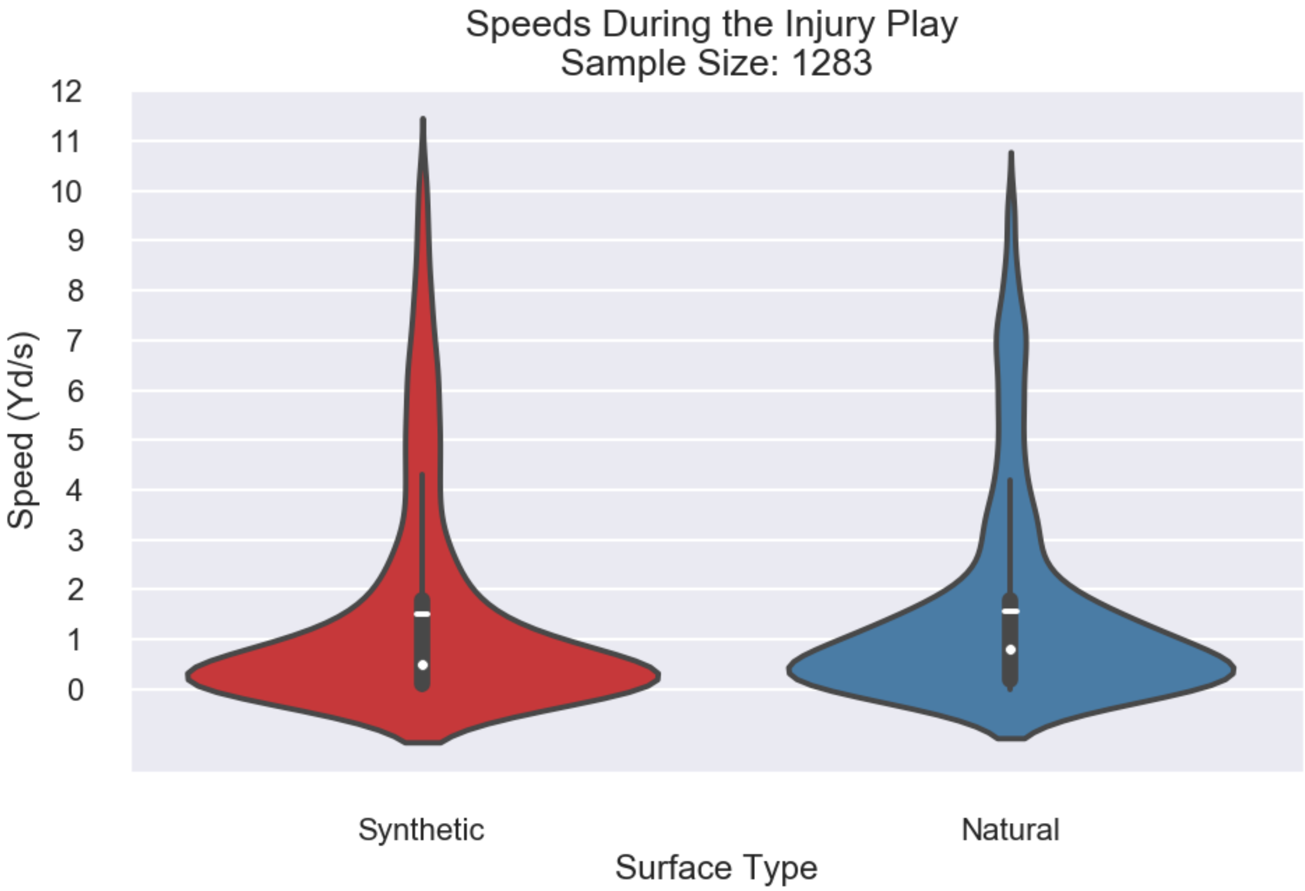
Injuries stem from these position groups

['RB' 'TE' 'Missing Data' 'DB' 'WR' 'OL' 'LB' 'DL']

- RB = Running back
- TE = Tight end
- DB = Defensive back
- WR = Wide Receiver
- OL = Offensive Line
- LB = Linebacker
- DL = Defensive Line

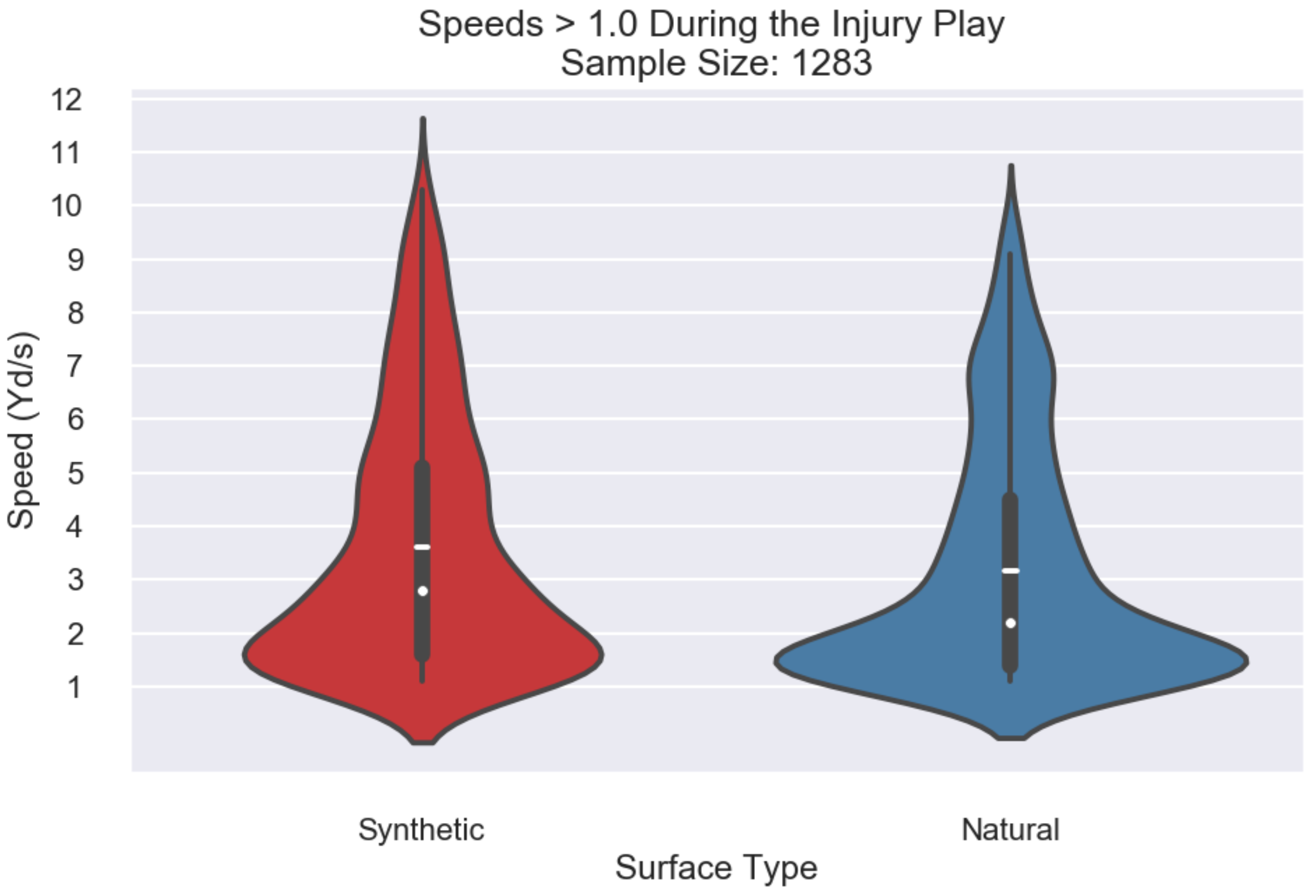
There are 14 entries of ‘Missing Data’

Understanding Player Movement

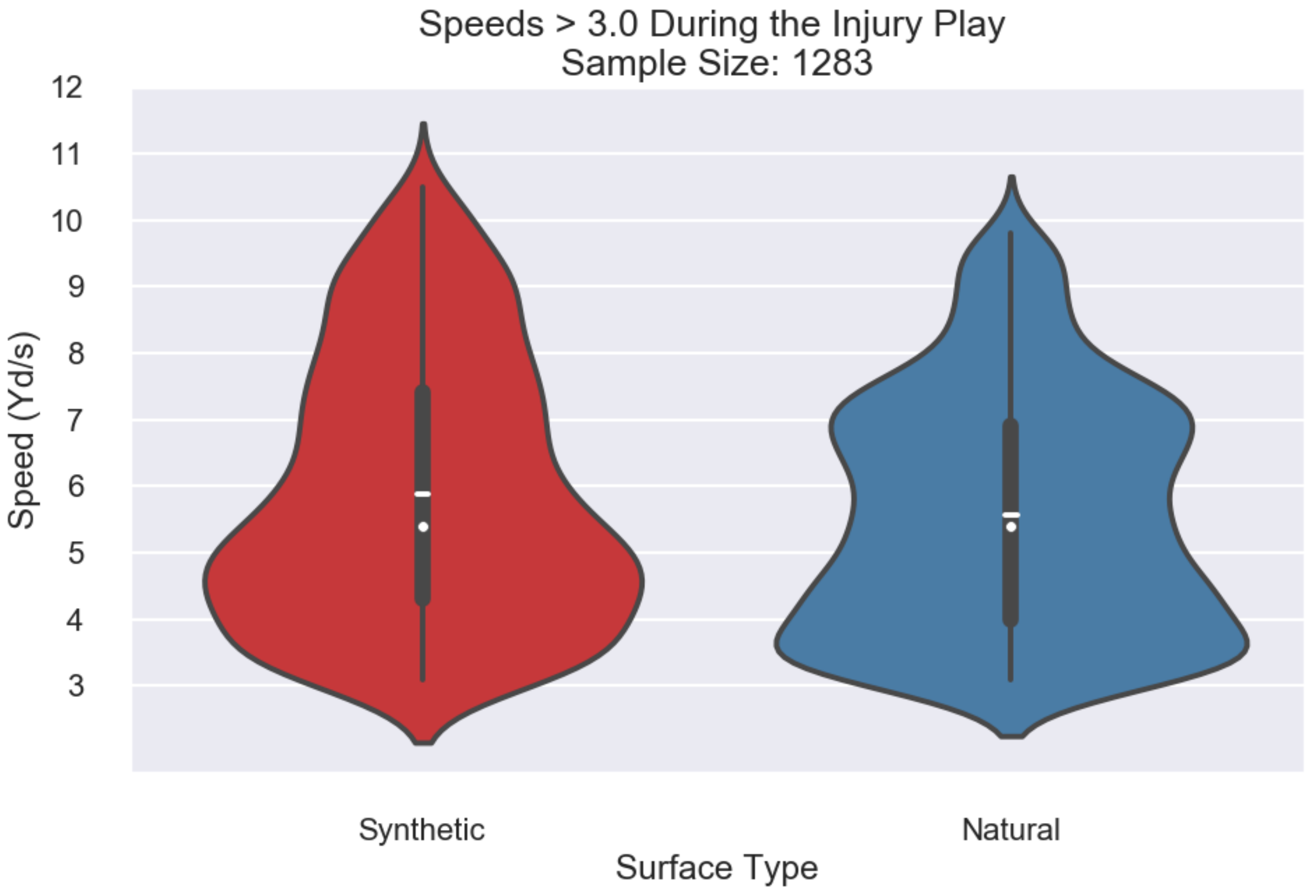


All Analysis is based off of the 77 entries that had a play key

- As you can see, a fair amount of the plays have speed at 0 for a lengthy period whilst the teams line up. If for example the RB isn't in motion until the snap this leads to a distribution that has a significant skew. Let's see what happens if we filter speeds if speed > 1 yard per second.
- I have also made the addition of a white line to represent the mean of each group
- Very important - these 2 groups have different population sizes: 4190 for synthetic and 4089 for natural. As a result, I will take an equal sample of both to accurately compare the 2 groups. I will base my sample on the larger population. With a confidence level of 99% and a confidence interval of +- 3, my sample will consist of 1283 individuals.



- Filtered with speeds > 1.0 you can still see quite a heavy skew. You can see, however, there is considerably more ‘low speeds’ say 2 yards / s or less on natural surfaces. Does this mean that acceleration is slower on natural vs synthetic? Let’s zoom further to speeds > 3.0



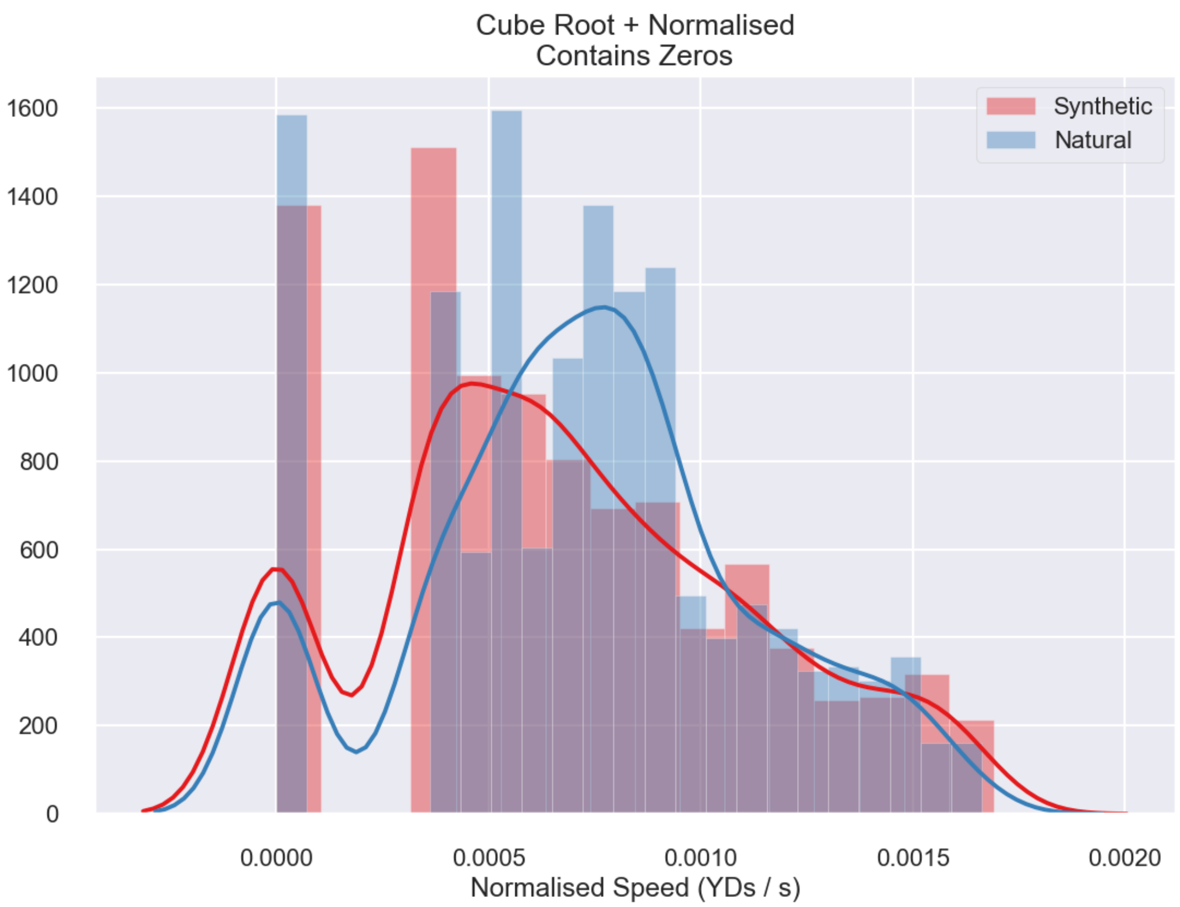
- Here are all the speeds over 3.0. Here, the median between synthetic and natural speeds are both 5.4. The mean on synthetic is consistently higher than natural pitches in all the violin plots so far.
- You can see that the natural pitches still have a higher proportion of speeds on the lower end – around 3 to 4. Interestingly, natural pitches have local maximum around 6.5 to 7.5 yards per second
- The maximum speed is higher on synthetic pitches

Important takeaways from injured players

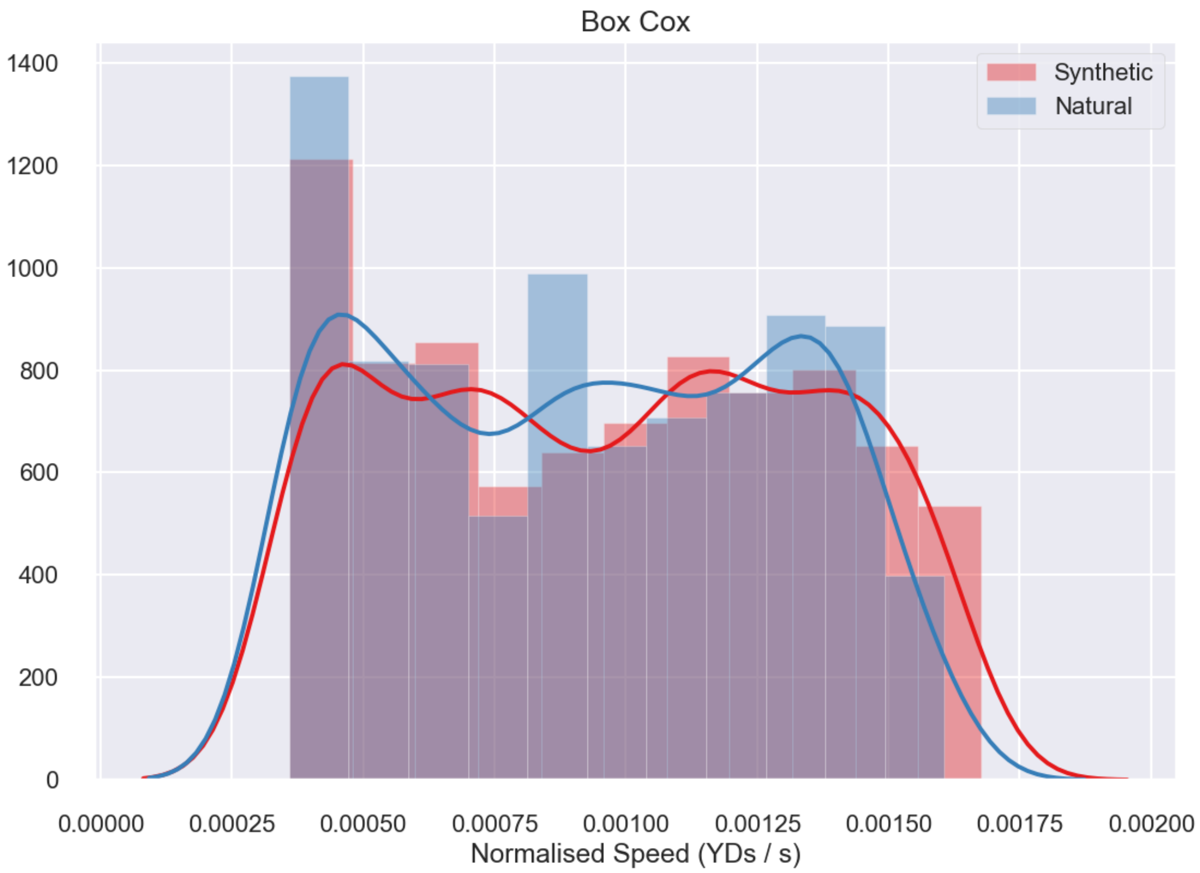
- Maximum speed on synthetic pitches was 10.5 yards / s
  - This is equivalent to 21.5 mph
- Maximum speed on natural pitches was 9.8 yards / s
  - Equivalent to 20 mph

Null: There is not a significant difference between the speeds on synthetics vs natural pitches

- The difficulty with this assessment is the heavy positive skew in the data



- As can be seen, a normal distribution is hard to achieve with such a high proportion of the data being 0 or around 0
- If we make the assumption that the average walking speed is 3.1 MPH or ~1.5 yards per second and we filter out all these values, lets see how it impacts the distribution. By reducing the population size and maintaining the sample size of 1283, this should be more than sufficient to be confident in the results



- Using Box-cox transformation and normalising, this was the distribution of data after filtering out the speeds under 1.5 (walking speed)
- Having used Bartlett's test, I identified that the variances between the synthetic and natural data was significantly different as such, I will use Welch's T-Test to identify any difference between the means
- My data will have
  - An identical population: 1283
  - A (relatively) normal distribution
  - Unequal variances; therefore, Welch's T Test
- *P*: 0.03 and thus is < 0.05
- This shows that there is a statistically significant difference between the speeds on the synthetic pitches and the natural ones

Is there a difference between the speed of the injury play versus the plays prior