

## 第3组 毛昊罡 周嘉莉 张翼

### 零 项目流程

#### 0.1 数据流程

#### 0.2 代码流程

### 一 业务理解

#### 1.1关于NFL

### 二 数据理解

#### 2.1 字段和键定义

##### 2.1.1 Injury Data File

##### 2.1.2 Play List

##### 2.1.3 Player Track Data

### 三 数据准备

#### 3.1 球场类型

#### 3.2 天气数据

### 四 建模

#### 4.1 数据可视化

##### 4.1.1 不同的场地类型与受伤次数

##### 4.2 受伤部位和场地存在一定关系

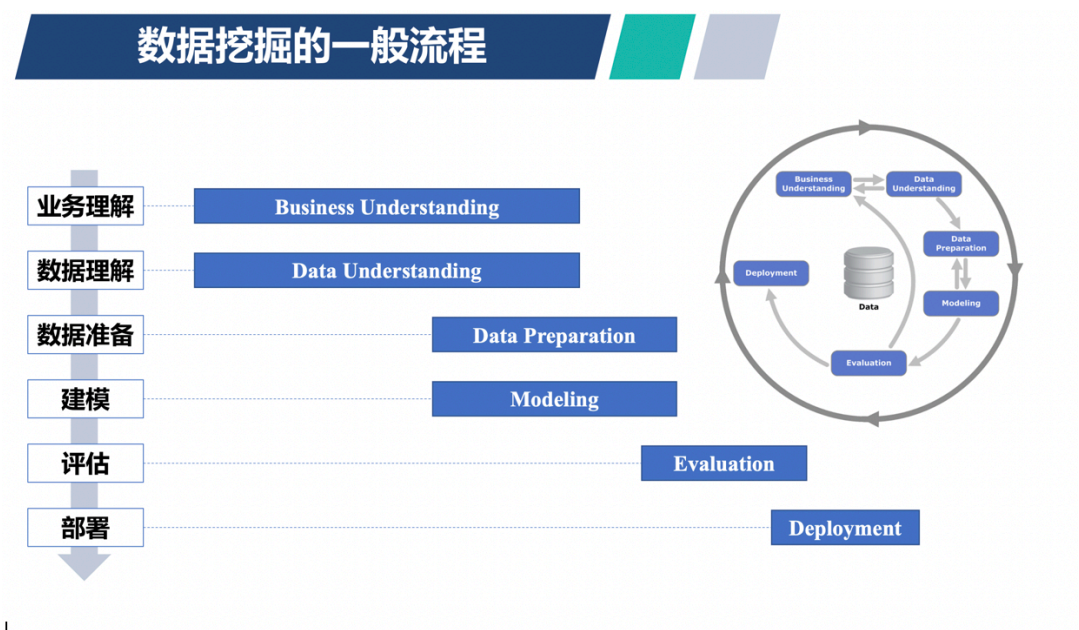
##### 4.3 平均速度统计

4.4 场地对运动员速度的影响

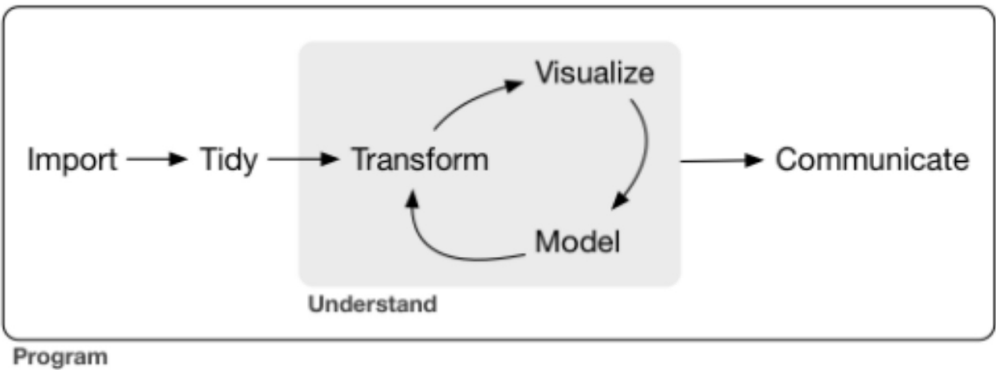
附 项目进度自查

零 项目流程

0.1 数据流程



0.2 代码流程



一 业务理解

## 1.1关于NFL

### 美国春晚

美国国家橄榄球联盟（National Football League）是美国最受欢迎的体育联盟，由32个专营权组成的联盟每年竞争赢得超级碗，这是世界上最大的年度体育赛事。NFL成立于1920年，为成功的现代体育联赛开发了该模式，包括国内和国际发行，广泛的收益共享，卓越的竞争以及遍布全国的强大特许经营权。

NFL致力于在运动相关伤害的诊断，预防和治疗方面取得进步。NFL正在进行的健康与安全工作包括对独立医学研究和工程进展的支持，以及致力于更好地保护玩家并使游戏更加安全的承诺，包括增强医疗协议以及改进我们的游戏教学和玩法。

从中学到的更多信息，联盟会评估和更改规则以发展比赛并尝试改善对球员的保护。仅从2002年以来，美国国家橄榄球联盟就进行了50项规则更改，旨在消除潜在的危险战术并减少受伤风险。

美国职业橄榄球大联盟居北美四大职业体育运动联盟之首，也是世界上规模最大的职业橄榄球大联盟。

联盟最早在1920年5月以美国职业橄榄球协会（American Professional Football Association）的名义成立，后来在1922年1月28日改名为国家橄榄球联盟。

NFL职业橄榄球联盟由32支来自美国不同地区和城市的球队组成，分为美国橄榄球联合会（AFC）和国家橄榄球联合会（NFC）两个联合会。每个联合会由东西南北四个分区组成，每个分区有四支球队和附属球员。

每年的赛事分为季前赛（Preseason）、常规赛（Regular Season）和季后赛（Playoff）三个时间段。季前赛共四周，作为常规赛的热身，每支球队在常规赛期间只会打16场比赛。

数据显示，2007年至今，英国人观看NFL电视转播的人数翻了一倍有余，到2018年，NFL官方的最新统计表明，NFL的英国粉丝数量达到了1400万。

#### 东部（NFC East）

[达拉斯牛仔队](#)（Dallas Cowboys）

[纽约巨人队](#)（New York Giants）

[费城老鹰队](#)（Philadelphia Eagles）

[华盛顿红皮队](#)（Washington Redskins）

#### 南部（NFC South）

[亚特兰大猎鹰队](#)（Atlanta Falcons）

卡罗莱纳黑豹队 (Carolina Panthers)  
新奥尔良圣徒队 (New Orleans Saints)  
坦帕湾海盗队 (Tampa Bay Buccaneers)

#### 西部 (NFC West)

亚利桑那红雀队 (Arizona Cardinals)  
旧金山49人队 (San Francisco 49ers)  
西雅图海鹰队 (Seattle Seahawks)  
洛杉矶公羊队 (Los Angeles Rams)

#### 北部 (NFC North)

芝加哥熊队 (Chicago Bears)  
底特律雄狮队 (Detroit Lions)  
绿湾包装工队 (Green Bay Packers)  
明尼苏达维京人队 (Minnesota Vikings)

## 二 数据理解

数据集中提供了三个文件，如下所述：

- Injury Record: .csv格式的伤害记录文件包含有关在两个赛季的常规赛期间发生的105下肢伤害的信息。可以使用PlayerKey，GameID和PlayKey字段将伤害链接到玩家历史记录中的特定记录。
- Play List: 有关比赛和比赛的详细信息包括球员分配的名册位置，体育场类型，场地类型，天气，比赛类型，比赛位置和位置组。
- Player Track Data:用于描述以10 Hz记录的播放过程中每个玩家的位置，方向，速度和方向（即，每秒记录10个观测值）。

### 2.1 字段和键定义

数据库	描述	球员键	比赛ID
受伤记录	关于受伤的信息	x	x
比赛文件	比赛、场地、球员-比赛水平的发挥信息	x	x
球员追踪	球员位置、速度、在一场比赛中球员个人的方向信息		

下面提供了数据集中包含的每个字段的描述及其相应的格式和描述。 **主键 外键** 变量以粗体指定。

### 2.1.1 Injury Data File

字段	格式	描述
球员关键 (PlayerKey)	XXXX	用五位数的数字键标识一位球员
比赛ID (Game ID)	PlayerKey-X	只识别球员的游戏 (不严格按照时间顺序)
发挥关键 (PlayKey)	PlayerKey-GameID-X	只识别比赛中球员的发挥 (按顺序)
身体部位 (BodyPart)	字符串	识别受伤的身体部位 (膝盖、脚踝、脚等……)
场地 (Surface)	字符串	识别受伤时的比赛场地 (天然的还是人造的)
DM_M1	1或0	One-hot编码, 指示由于受伤错过了1天或更多天
DM_M7	1或0	One-hot编码, 指示由于受伤错过了7天或更多天
DM_M28	1或0	One-hot编码, 指示由于受伤错过了28天或更多天
DM_M42	1或0	One-hot编码, 指示由于受伤错过了42天或更多天

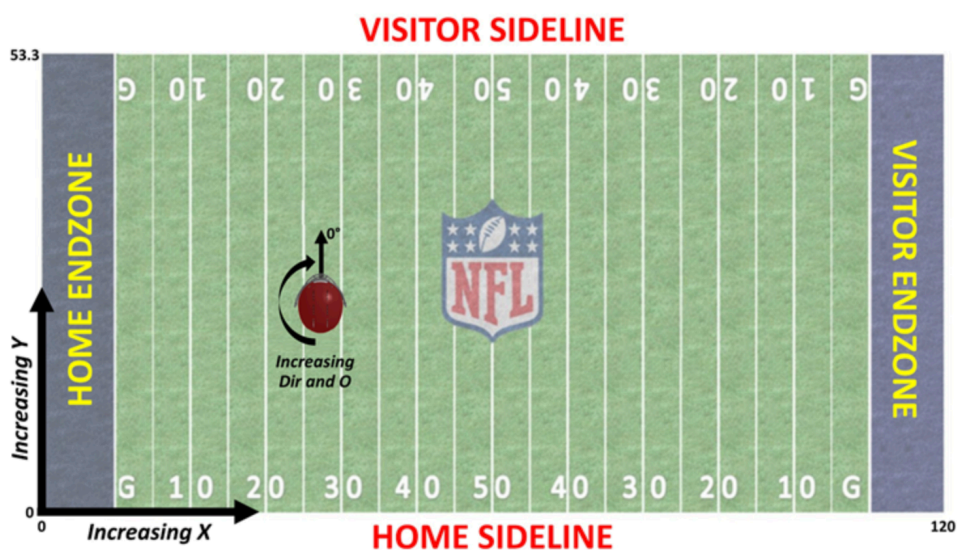
### 2.1.2 Play List

字段	格式	描述
球员关键 (PlayerKey)	XXXXX	用五位数的数字键标识一位球员
比赛ID (Game ID)	PlayerKey-X	只识别球员-比赛 (该索引不严格按照时间顺序)-识别
发挥关键 (PlayKey)	PlayerKey-GameID-X	只识别比赛中球员的发挥 (按比赛顺序)
名册位置(RosterPosition)	字符串	提供球员的名册位置 (即返回)
*比赛时间(PlayerDay)	整型	一个整型序列, 它反映了球员在比赛中的参与时间 球员参与的顺序。
*球员比赛 (PlayerGame)	整型	只识别一个球员的比赛; 匹配比赛ID的最后一个整 赛的时间顺序)
体育场类型(StadiumType)	字符串	体育馆类型的自由文字说明 (闭式、开放式圆顶等
场地类型(FiledType)	字符串	场地种类的分类描述 (自然的还是人造的)
温度(Temperature)	浮点数	比赛开始时的现场温度 (并不总是可用的-对于闭 于受温度和天气的影响, 此项可能是不相关的)
天气 (Weather)	字符串	描述体育场天气的自由文字说明 (由于受温度和天 是不相关的)
玩法类型 (PlayType)	字符串	玩法的分类描述 (传球、跑、踢等等)
PlayerGamePlay		指示球员在比赛过程中参与次数的顺序索引
位置 (Position)	字符串	指示球员在比赛中的位置的分类变量 (RB,QB,DE 不同
队伍位置 (Position Group)	字符串	表示球员在比赛中所处位置的位置组的分类变量

### 2.1.3 Player Track Data

字段	格式	描述	补充
发挥关键(PlayKey)	字符串		
时间 (time)	浮点型	从该比赛的NGS轨道开始的秒数（浮点数）	这是
事件 (event)	字符串	随时间变化的比赛细节（字符串）（huddle brea	
x	数字列表	沿着场地的长轴随时间变化的球员位置	0~12
y	数字列表	沿着场地的短轴随时间变化的球员位置	0~53
dis	数字列表	从先前时间点到时间索引的行进距离	距离
s	数字列表	时间索引上该特定时间点的估计速度	码每
o	数字	方向-球员面朝的角度（deg）	0~360
dir	数字	方向-球员运动的角度（deg）	0~360

处理玩家轨迹数据时，建议使用x，y位置数据计算速度，并将这些计算出的速度用于任何分析（尽管我们提供了NGS系统报告的速度变量）。x和y坐标的原点定义为起始端区和起始边线的角（见下图）。由方向和方向定义的角度是从坐标系的y轴参考的。



注意，方向变量不应被视为玩家面对的实际方向的可靠指标。这项研究的记录来自NFL的多个赛季，在此期间，使用了不同的系统来计算和记录球员的方位。在一个游戏内以及游戏中的多个游戏中，单个玩家的方向将被一致地记录，但是报告的方向的“地理位置”（即玩家所面对的实际方向）可能与方向变量的“地理位置”不匹配在同一戏剧中。对于参与多个季节的那些球员，用于记录方向变量的地理位置在各个季节之间都不相同。方向变量可用于表征玩家在比赛过程中转动或旋转多少。

## 三 数据准备

清洗数据(错别字/空字段)

### 3.1 球场类型

体育馆类型	count
	16,910
碗型	465
封闭的圆顶	1,011
多云的	178
圆顶	9,376
圆顶， 封闭	1,059
半球形	985
圆顶， 封闭	3,076
圆顶， 开放	1,779
半球形， 开放式	807
亨氏球场	389
室内	6,892



室内，开放式屋顶	479
室内，屋顶封闭	547
在室内	22,805
打开	4,124
Outdoor	914
外门	778
优胜者	595
户外	145,032
户外零售屋顶开放	601
户外活动	32,956
Outdoor	356
外	341
零售 屋顶-封闭	2,235
零售 屋顶-开放	486
零售 屋顶关闭	414
零售 封闭式	2,015
零售 屋顶开放	486
伸缩屋顶	8,914

分为五种类型

- outdoor
- indoor\_closed

- indoor\_open
- dome\_closed
- dome\_open

## 3.2 天气数据

天气	count
	18,691.0
降雨几率10%	413.0
30%的降雨机会	354.0
明确	25,923.0
晴冷	758.0
清爽	638.0
晴朗晴朗	488.0
晴朗晴朗	613.0
清澈温暖	1,252.0
晴朗的天空	812.0
晴朗的天空	842.0
晴转多云	495.0
多云的	987.0
多云的	60,705.0
阴冷	745.0
阴凉	835.0

可能有雷雨。风向WNW，时速10-20英里。	246.0
多云，雨量变化50%	769.0
可能有雨	456.0
多云，第二季度开始起雾	766.0
多云，积雪1-3“”	616.0
阵雨	384.0
冷	549.0
受控气候	6,355.0
库迪	278.0
公平	2,218.0
朦胧	1,043.0
耐热指数95	455.0
大湖效应雪	383.0
室内	3,038.0
在室内	3,460.0
小雨	1,855.0
大部分多云	2,389.0
大部分多云	11,176.0
主要是库迪	601.0
大多时间晴天	1,090.0
大多时间晴天	3,180.0

大多时间晴八	3,100.0
晴间多云	414.0
不适用（室内）	5,517.0
室内不适用	1,906.0
灰蒙蒙	865.0
部分清晰	746.0
局部阴天	1,311.0
局部阴天	28,463.0
局部性	302.0
局部晴朗	517.0
局部晴朗	2,493.0
派对多云	230.0
雨	8,349.0
降雨几率40%	691.0
可能会下雨，气温在40s以下。	553.0
花洒淋浴	464.0
多雨的	499.0
零星阵雨	615.0
阵雨	516.0
雪	946.0
太阳与云	385.0

阳光明媚	51,728.0
阳光明媚	698.0
晴天和寒冷	535.0
阳光明媚，温暖	687.0
晴朗的天空	618.0
晴天，最高到80年代	632.0
阳光明媚，刮风	467.0

分为五个组：

- rain
- overcast
- clear
- snow
- none

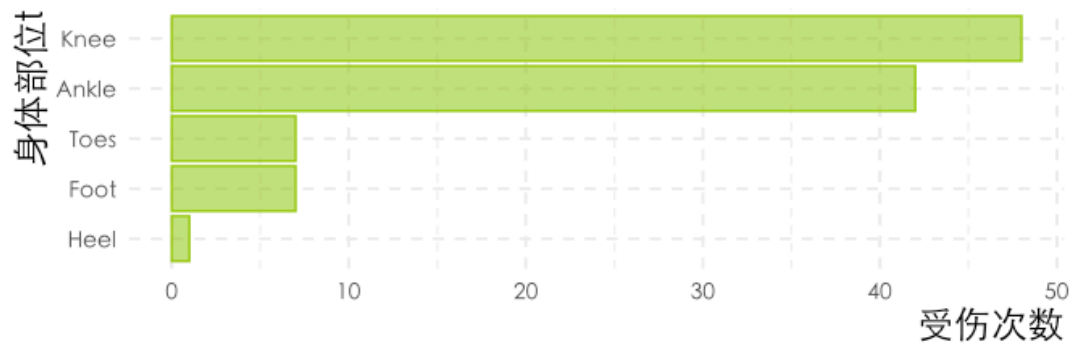
## 四 建模

### 4.1 数据可视化

#### 4.1.1 不同的场地类型与受伤次数

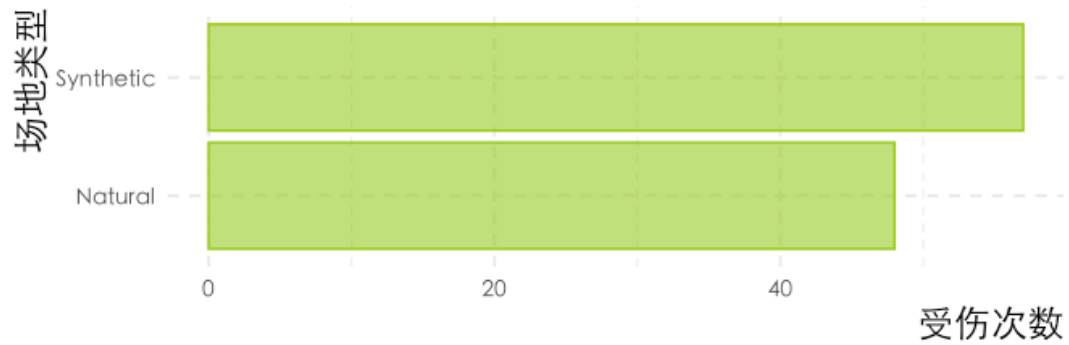
## NFL运动员各身体部位受伤次数

86%的受伤发生在是由膝盖和脚踝



## 不同的场地类型与受伤次数

人造场地的伤害要比自然场地稍微多一些

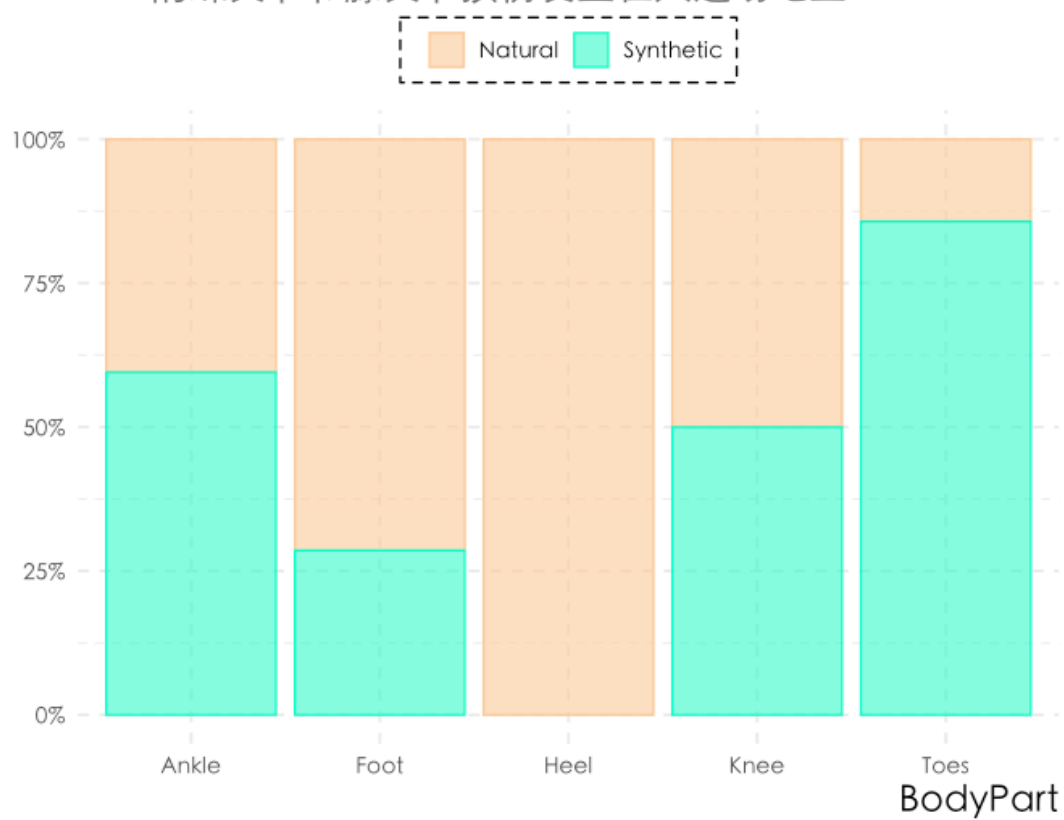


### 4.2 受伤部位和场地存在一定关系

# Heel关节只被记录了一次

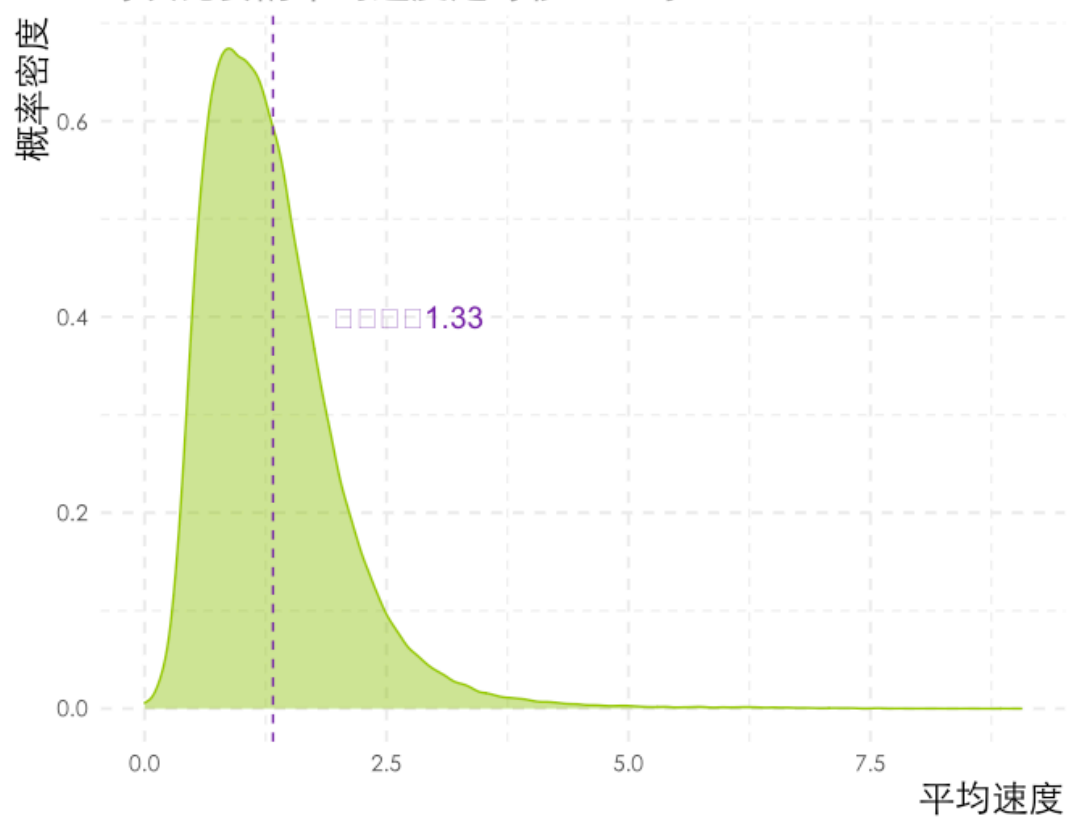
## 受伤部位和场地存在一定关系

60%的踝关节和膝关节损伤发生在人造场地上



### 4.3 平均速度统计

NFL运动员的平均速度稍有偏差  
每次比赛的平均速度是每秒1.33码

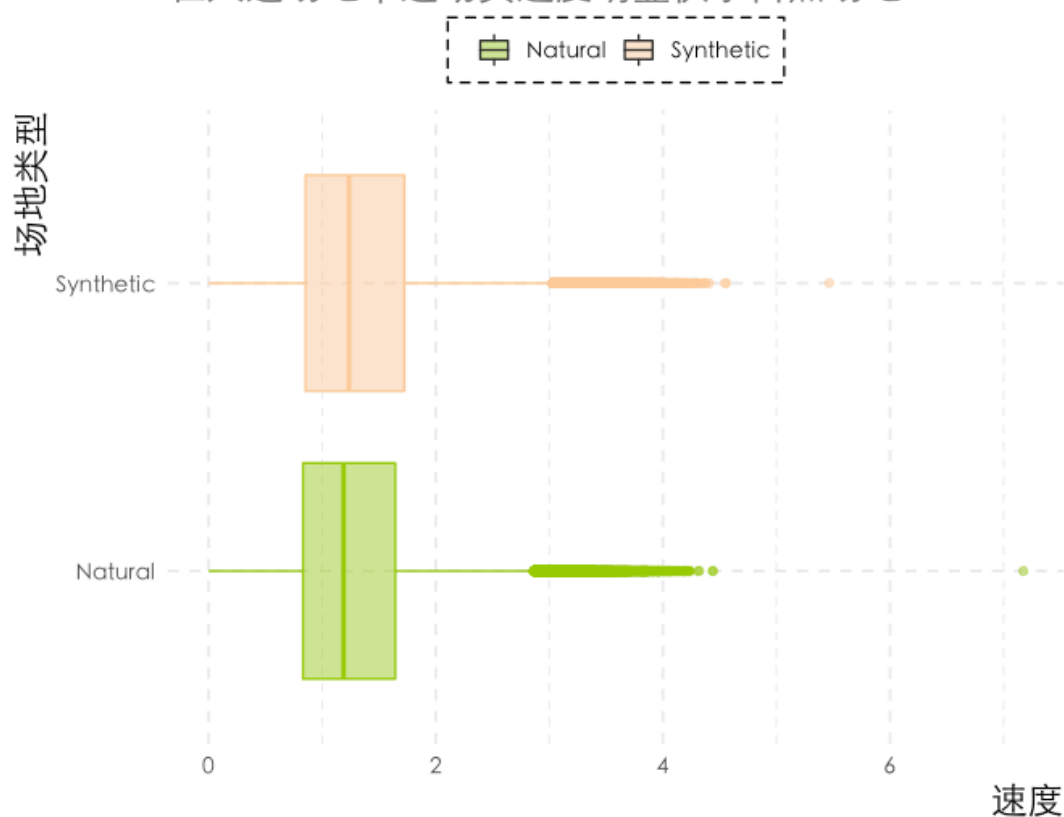


#### 4.4 场地对运动员速度的影响



## 场地对运动员速度的影响

在人造场地中运动员速度明显快于自然场地



## 附1 项目进度自查

项目地址<https://github.com/AldonahZero/nfl-playing-surface-analytics>

- ☒ Import 包环境依赖,导入数据
- ☒ Tidy 清洗数据(错别字/空字段)
- ☒ Transform数据转换/翻译工作
- ☒ vis 数据可视化
- ☐ model 数据建模
- ☐ com 评估工作
- ☐ ppt制作 讲稿编写 排练

## 附2 数据解释

常用数据字典对照	解释
Ankle	脚踝
Foot	脚背
Hell	脚后跟
Knee	膝盖
Toes	脚趾