

Explora tu equipo de fantasía de atletismo

Estadística computacional

31/1/2022

1) El atletismo necesita una nueva raza de scouts y managers

El atletismo se remonta a los Juegos Olímpicos originales. Desde entonces, poco ha cambiado. Los atletas compiten como individuos, buscando lanzar lo más lejos, saltar lo más lejos (o lo más alto) y correr lo más rápido. Pero a la gente le gusta animar a los equipos, ondear pancartas y gritar como locos durante los partidos, llevar las camisetas de sus jugadores favoritos y mantenerse fiel a su equipo en las buenas y en las malas.

¿Y si el atletismo fuera un deporte de equipo? Podría ser potencialmente más interesante y nos daría un nuevo conjunto de análisis deportivos que discutir. Incluso podríamos reducir los incentivos para hacer cosas desagradables en la búsqueda de altius, fortius y citius.

Este conjunto de datos contiene resultados de atletas estadounidenses en los saltos horizontales (triple salto y salto de longitud) y los lanzamientos (lanzamiento de peso, disco, jabalina, martillo y peso). Leamos esto y examinemos la jabalina femenina.

Listado de Atletas

```
## # A tibble: 68 x 1
##   Athlete
##   <chr>
## 1 Abigail Gomez
## 2 Abigail Gomez Hernandez
## 3 Alicia DeShasier
## 4 Allison Updike
## 5 Alyssa Olin
## 6 Ariana Ince
## 7 Asha Bramwell
## 8 Asia Easley
## 9 Avione Allgood
## 10 Bethany Drake
## # ... with 58 more rows
```

Eventos realizado en esta categoría

```
## # A tibble: 26 x 1
##   EventID
##   <dbl>
## 1      8
## 2     20
## 3    218
```

```
## 4      238
## 5      180
## 6      247
## 7      178
## 8      174
## 9      172
## 10     173
## # ... with 16 more rows
```

Las 5 Atletas que mas han participado en eventos

```
## # A tibble: 5 x 2
## # Groups:   Athlete [5]
##   Athlete      n
##   <chr>      <int>
## 1 Ariana Ince      13
## 2 Leigh Petranoff  12
## 3 Kimberley Hamilton 11
## 4 Brittany Borman   9
## 5 Melissa Fraser    8
```

2) Los managers adoran los datos ordenados

Esta vista muestra los resultados de cada atleta en las competencias individuales de atletismo. Los atletas tienen seis lanzamientos, pero en estos encuentros sólo importa uno, el más largo. Si todo lo que quisiéramos hacer fuera hablar de atletismo normal, tendríamos una tarea muy fácil: crear una nueva columna tomando el máximo de cada fila, organizar el marco de datos por esa columna en orden descendente y ya estaría.

Pero nuestros managers tienen que hacer y saber mucho más que eso. Este es un deporte de estrategia, en el que cada lanzamiento importa. Los directivos necesitan un análisis más profundo para elegir sus equipos, elaborar su plan y tomar decisiones el día del partido.

Primero tenemos que hacer que esta vista estándar “amplia” ordene los datos. No hemos terminado del todo con la vista amplia, pero los datos ordenados nos permitirán calcular nuestras estadísticas resumidas.

```
## 'summarise()' has grouped output by 'Athlete'. You can override using the '.groups' argument.
```

```
## # A tibble: 6 x 5
## # Groups:   Athlete [3]
##   Athlete      EventID maximo Lanzamiento Valor
##   <chr>      <dbl>   <dbl> <chr>      <dbl>
## 1 Kara Patterson      747    64.9 1         64.9
## 2 Kara Winger      1859    62.9 1         56.9
## 3 Kara Winger      1740    62.8 1         62.8
## 4 Kara Patterson      20    62.4 1         56.7
## 5 Brittany Borman      20    62.0 1          58
## 6 Brittany Borman      747    61.8 1        61.1
```

```
## Selecting by Valor
```

```
## # A tibble: 584 x 5
## # Groups:   Athlete [68]
##   Athlete      EventID maximo Lanzamiento Valor
```

```
##      <chr>          <dbl> <dbl> <chr>          <dbl>
## 1 Kara Patterson      747   64.9 1             64.9
## 2 Kara Patterson      747   64.9 6             64.4
## 3 Kara Winger        1859   62.9 6             62.9
## 4 Kara Winger        1740   62.8 1             62.8
## 5 Kara Patterson      20    62.4 3             62.4
## 6 Kara Patterson      20    62.4 6             62.3
## 7 Brittany Borman     20    62.0 3             62.0
## 8 Kara Winger        1740   62.8 6             61.9
## 9 Brittany Borman     747   61.8 6             61.8
## 10 Kara Winger       1740   62.8 4             61.6
## # ... with 574 more rows
```

3) Cada lanzamiento es importante

Un lanzamiento es una falta si el atleta comete una infracción técnica durante el lanzamiento. En jabalina, la falta más común es sobrepasar la línea de lanzamiento. Tradicionalmente, el lanzamiento se puntúa como “F” y no tiene más importancia. Los atletas también pueden optar por pasar de un lanzamiento, que se puntúa con una “P”, si están contentos con sus lanzamientos anteriores y quieren “reservarse” para más adelante.

¿Recuerdas cuando decíamos que cada lanzamiento es importante? Aquí, el objetivo no es que cada jugador tenga un gran lanzamiento. Todos sus lanzamientos en cada prueba se suman, y el equipo con la mayor distancia total gana el punto. Las faltas se puntúan como 0 y los pases, bueno, su director y sus compañeros de equipo no estarían contentos.

Aquí examinamos qué atletas cubren más distancia en cada uno de sus encuentros, junto con dos formas de hablar de su consistencia.

```
## ‘summarise()’ has grouped output by ‘Athlete’. You can override using the ‘.groups’ argument.
```

```
## # A tibble: 10 x 6
## # Groups:   Athlete [4]
##   Athlete          EventID dis_total promedio_lanz      sd Validos
##   <chr>          <dbl>     <dbl>         <dbl> <dbl>    <int>
## 1 Abigail Gomez      238      152.         25.3 27.7      3
## 2 Abigail Gomez      498      244.         40.7 20.0      5
## 3 Abigail Gomez      511      207.         34.5 26.8      4
## 4 Abigail Gomez     1566      222.         37.0 28.7      4
## 5 Abigail Gomez     1575      155.         25.8 28.3      3
## 6 Abigail Gomez Hernandez 1727      135.         22.5 24.7      3
## 7 Alicia DeShasier    178      270.         45.0 22.1      5
## 8 Alicia DeShasier    247      320.         53.3  2.26      6
## 9 Alicia DeShasier    938      275.         45.8 22.5      5
## 10 Allison Updike     681      147.         24.5 26.9      3
```

4) Encuentre a los que se quedan en la cima de la clasificación

En muchas competiciones de atletismo tradicionales, después de los tres primeros lanzamientos, los líderes del campo se reducen a los ocho mejores atletas (a veces más, a veces menos). Al igual que en la general, esto se basa únicamente en el mejor lanzamiento de esos tres primeros.

Damos la elección a los managers. De los tres atletas que comienzan cada prueba, el seleccionador elige a los dos que continuarán en los tres últimos lanzamientos. El seleccionador tendrá que saber qué jugadores

tienden a animarse -o al menos a mantener su forma- en las últimas fases de un partido. También necesita saber si los tres primeros lanzamientos de un jugador son coherentes con su historial de juego. De lo contrario, podrían tomar una mala decisión sobre quién se queda basándose únicamente en la muestra que se está desarrollando delante de ellos.

Por ahora, examinemos sólo nuestra estadística principal -la distancia total recorrida- para ver las diferencias entre las fases iniciales y finales del partido.

```
## 'summarise()' has grouped output by 'Athlete'. You can override using the '.groups' argument.
```

```
## # A tibble: 10 x 5
## # Groups:   Athlete [5]
##   Athlete      EventID fase_inicial fase_final diferencia
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Sarah Firestone    1703      150.      50.3     -99.7
## 2 Shana Woods        174       30.9      30.8     -0.0500
## 3 Sharon Day-Monroe   498      128.      77.6     -50.4
## 4 Sharon Day-Monroe   511      85.4      84.4      -1.05
## 5 Sharon Day-Monroe  1766      98.0     148.      50.5
## 6 Sharon Day-Monroe  1773     143.      84.0     -59.1
## 7 Tairyn Montgomery   176      93.4     142.      48.2
## 8 Tiffany Forbito     498     143.      89.7     -53.4
## 9 Tiffany Forbito     511     98.8      45.1     -53.7
## 10 Tiffany Forbito    681     137.      45.5     -91.2
```

5) Unir las piezas para una nueva mirada a los atletas

Las estadísticas agregadas están en dos marcos de datos. Uniéndolos dos, podemos echar un primer vistazo a la comparación de los atletas.

```
## # A tibble: 10 x 5
##   Athlete      dis_total sd validos diferencia
##   <chr>      <dbl> <dbl> <dbl>      <dbl>
## 1 Abigail Gomez    207.  27.7     4      -48
## 2 Abigail Gomez Hernandez 135.  24.7     3      45.6
## 3 Alicia DeShasier   275.  22.1     5      53.5
## 4 Allison Updike    147.  26.9     3     -46.6
## 5 Alyssa Olin       203.  26.2     4     -49.3
## 6 Ariana Ince       269.  22.3     5      -4.57
## 7 Asha Bramwell     126.  16.4     4      56.1
## 8 Asia Easley       188.  19.7     4.5     64.1
## 9 Avione Allgood    217.  28.0     4       -5
## 10 Bethany Drake     195.  25.2     4     -1.81
```

6) Normalizar los datos para comparar las estadísticas

Las cuatro estadísticas de resumen -distancia total, desviación estándar, número de lanzamientos exitosos y nuestra medida de temprano vs. tarde- están en diferentes escalas y miden cosas muy diferentes. Los directivos deben ser capaces de compararlas entre sí y luego sopesarlas en función de lo que sea más importante para su visión y estrategia para el equipo. Una simple normalización permitirá estas comparaciones.

```
##           Athlete TotalDistance StandardDev Success      diff
## 1      Kara Winger      1.0000000  0.05728290      1.00 0.4808205
## 2 Dominique Ouellette  0.8507808  0.03901900      1.00 0.4885542
## 3      Haley Crouser   0.8442649  0.01845056      1.00 0.4972397
## 4      Maggie Malone   0.8322025  0.74411139      0.75 0.7657529
## 5      Heather Bergmann 0.7961228  0.01789325      1.00 0.4863411
## 6      Madalaine Stulce 0.7649973  0.09112837      1.00 0.4443175
```

7) ¿Qué es lo que más importa a la hora de confeccionar la plantilla?

Los managers tienen que decidir qué tipo de jugadores quieren en su equipo: quiénes coinciden con su visión, quiénes tienen las habilidades que necesitan para practicar su estilo de atletismo y, en última instancia, quiénes conseguirán las victorias. Un entrenador con aversión al riesgo querrá jugadores que rara vez cometan faltas. El entrenador con ojos de acero querrá a los jugadores que puedan conseguir la victoria con sus lanzamientos finales.

Como en cualquier otro deporte (o profesión), rara vez un jugador será igual de fuerte en todas las áreas. Los managers tienen que hacer concesiones a la hora de seleccionar sus equipos. Nuestra primera tanda de entrenadores tiene la desventaja añadida de seleccionar a los jugadores basándose en los datos de un deporte relacionado pero distinto. Nuestros datos proceden de competiciones tradicionales de atletismo, donde las motivaciones y los objetivos son muy diferentes a los nuestros.

Por eso los managers ganan mucho dinero y son despedidos cuando los resultados se van al garete.

```
## ¿Desea ponderar de forma manual? Elija S para Sí y N para No):
```

```
## [1] "La ponderacion sera aleatoria"
```

```
## Peso de Equipo
```

```
##      Home: Distancia: 37 Variabilidad: 38 Exitos: 18 Diferencia: 7
```

```
##           Athlete TotalDistance StandardDev Success      diff seleccion
## 1      Maggie Malone   0.8322025  0.7441114   0.750 0.7657529  77.92799
## 2 Alicia DeShasier    0.7647460  0.6810183   0.750 0.7293451  72.77972
## 3      Kara Patterson  0.7043798  0.7964048   0.625 0.4665667  70.84140
## 4      Nicolle Murphy  0.5618740  0.8828093   0.500 1.0000000  70.33609
## 5      Ariana Ince     0.7442829  0.6854711   0.750 0.4531220  70.25822
```

8) Conozca a sus jugadores

¡Los datos han hablado! Ahora tenemos a nuestros cinco lanzadores de jabalina, pero todavía no los conocemos realmente. El marco de datos **javelin_totals** tiene los datos que se utilizaron en el proceso de decisión, así que los sacaremos. Esto nos da una idea de lo que cada uno aporta al equipo.

También podemos ver cómo se comparan con el grupo de atletas del que partimos tomando la media y el máximo de cada estadística.

9) Defiende tus argumentos ante la dirección del equipo

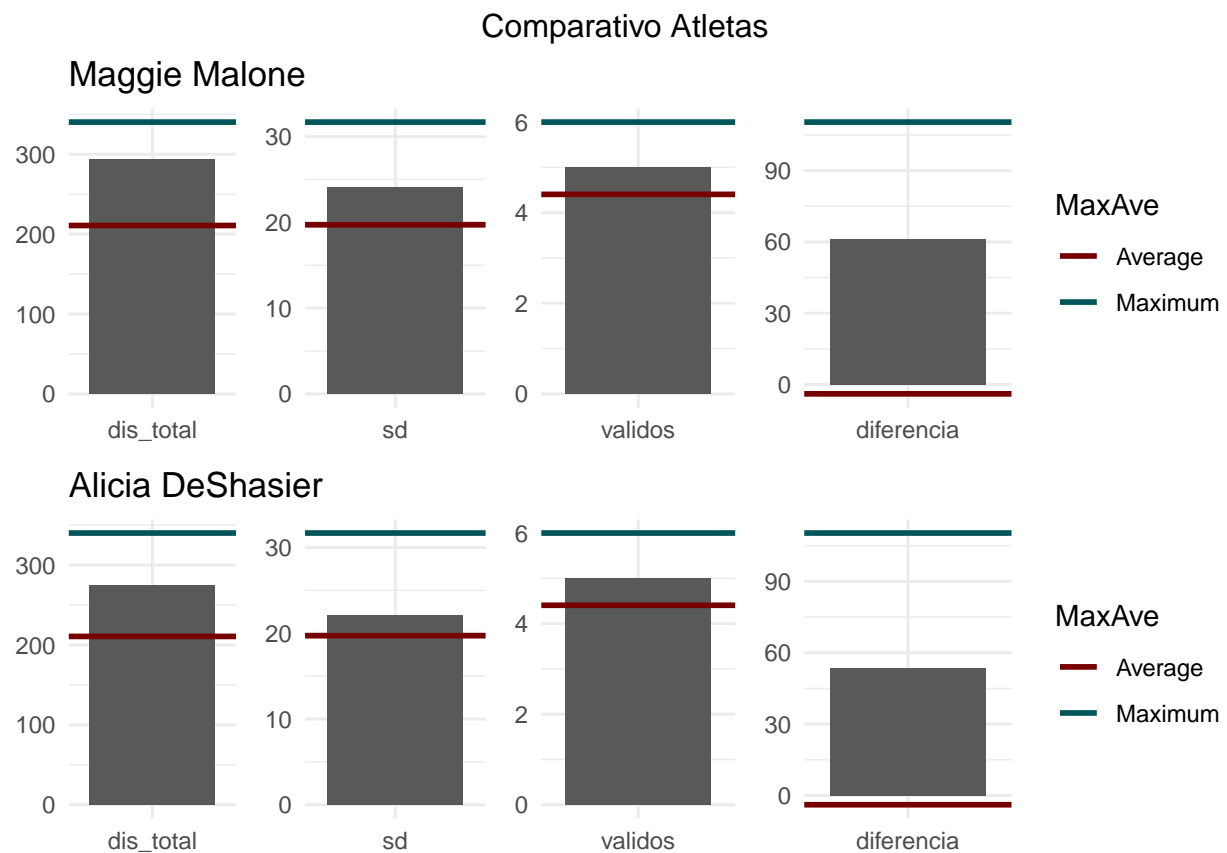
El manager sabe lo que quiere del equipo y tiene los datos que respaldan sus decisiones, pero aún así tiene que defender sus decisiones ante los propietarios del equipo. Al fin y al cabo, son ellos los que extienden los cheques.

Los propietarios son personas ocupadas. Muchos de ellos tienen otros trabajos y otras empresas. Confían en sus directivos, así que mientras el directivo pueda darles una presentación visual fácil de entender de por qué deberían fichar a estos cinco atletas de entre todos los demás, lo aprobarán.

Una serie de gráficos que muestren cómo se compara cada atleta con el máximo y la media de cada estadística será suficiente para ellos.

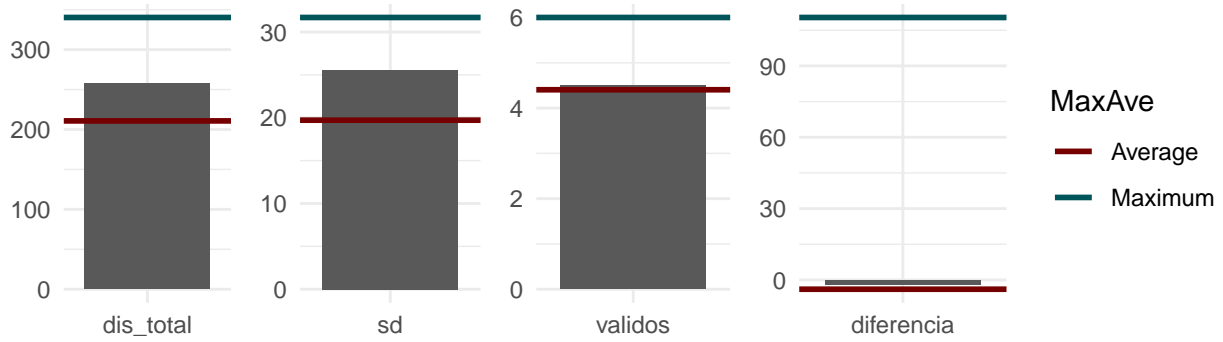
```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
## combine
```

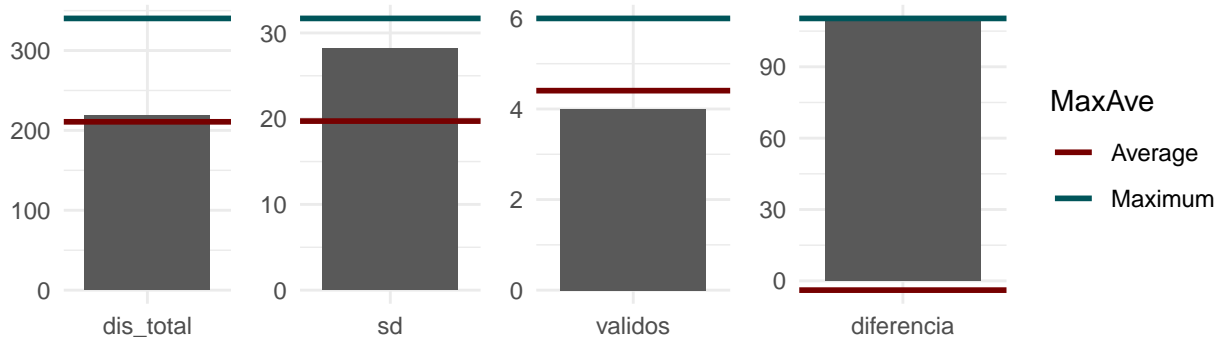


Comparativo Atletas

Kara Patterson

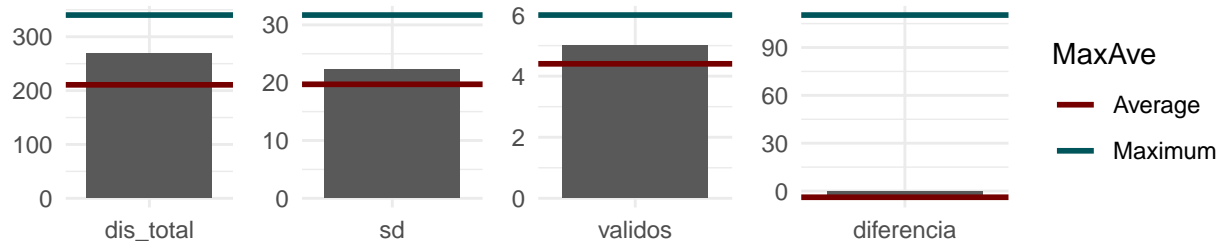


Nicolle Murphy

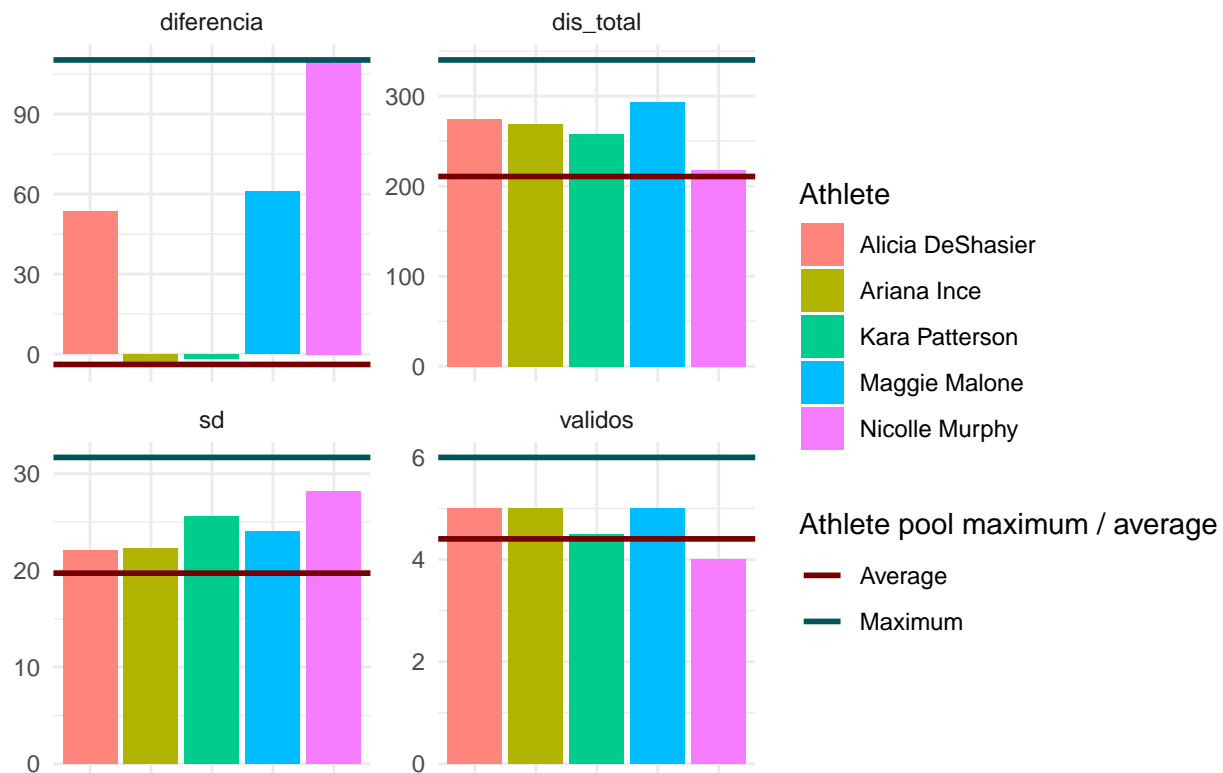


Comparativo Atletas

Ariana Ince



AnyTown Athletic Club: Women's Javelin



/newpage

10) Tiempo de lanzamiento

Antes de que comience la temporada de atletismo, el director realizará análisis similares para los demás lanzamientos, los saltos y las pruebas de carrera. Luego jugará con diferentes permutaciones de su equipo y de su rival para llegar a la mejor alineación y tomar las mejores decisiones el día del partido. De momento, como es lo que mejor conocemos y casi no tenemos tiempo, vamos a simular un simple partido de jabalina.

El ganador es el equipo que lanza la mayor distancia combinada: seis lanzamientos de cada uno de tus tres jugadores contra seis lanzamientos de cada uno de los tres jugadores del adversario.

```
## ¿Desea ponderar de forma manual? Elija S para Sí y N para No):
```

```
## [1] "La ponderacion sera aleatoria"
```

```
## Peso de Equipo
```

```
## Home: Distancia: 65 Variabilidad: 24 Exitos: 9 Diferencia: 2
```

```
## ¿Desea ponderar de forma manual? Elija S para Sí y N para No):
```

```
## [1] "La ponderacion sera aleatoria"
```

```
## Peso de Equipo
```

```
## Home: Distancia: 6 Variabilidad: 77 Exitos: 12 Diferencia: 5
```

##	Athlete	TotalDistance	StandardDev	Success	diff	seleccion
## 5	Krista Woodward	0.7427033	0.6761445	0.75	0.2679421	71.78907
## 3	Alicia DeShasier	0.7647460	0.6810183	0.75	0.7293451	74.26162
## 4	Ariana Ince	0.7442829	0.6854711	0.75	0.4531220	72.48594

##	Athlete	TotalDistance	StandardDev	Success	diff	seleccion
## 4	Avione Allgood	0.5571352	0.8780375	0.5	0.4510756	79.20708
## 5	Hannah Carson	0.4691438	0.8602909	0.5	0.4579764	77.34715
## 2	Nicolle Murphy	0.5618740	0.8828093	0.5	1.0000000	82.34756

[1] "Javelin match, Final Score: Home: 624.64 - Away: 483.8"

[1] "Win!"

[1] "Win!"