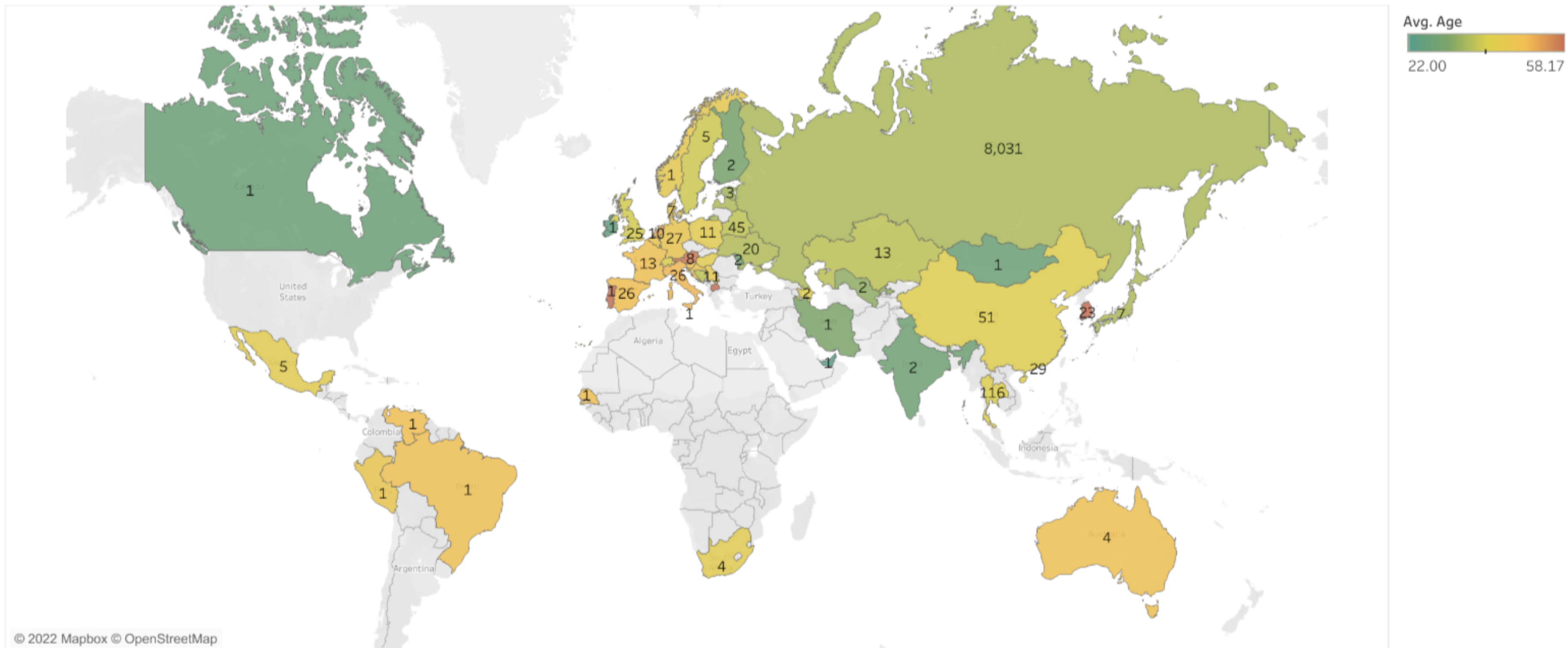


EL MARATÓN DE MOSCÚ 🇷🇺

Pronto correré mi primer maratón y me dí a la tarea de analizar datos de un maratón pasado que puedan darme algunos tips. Veamos...

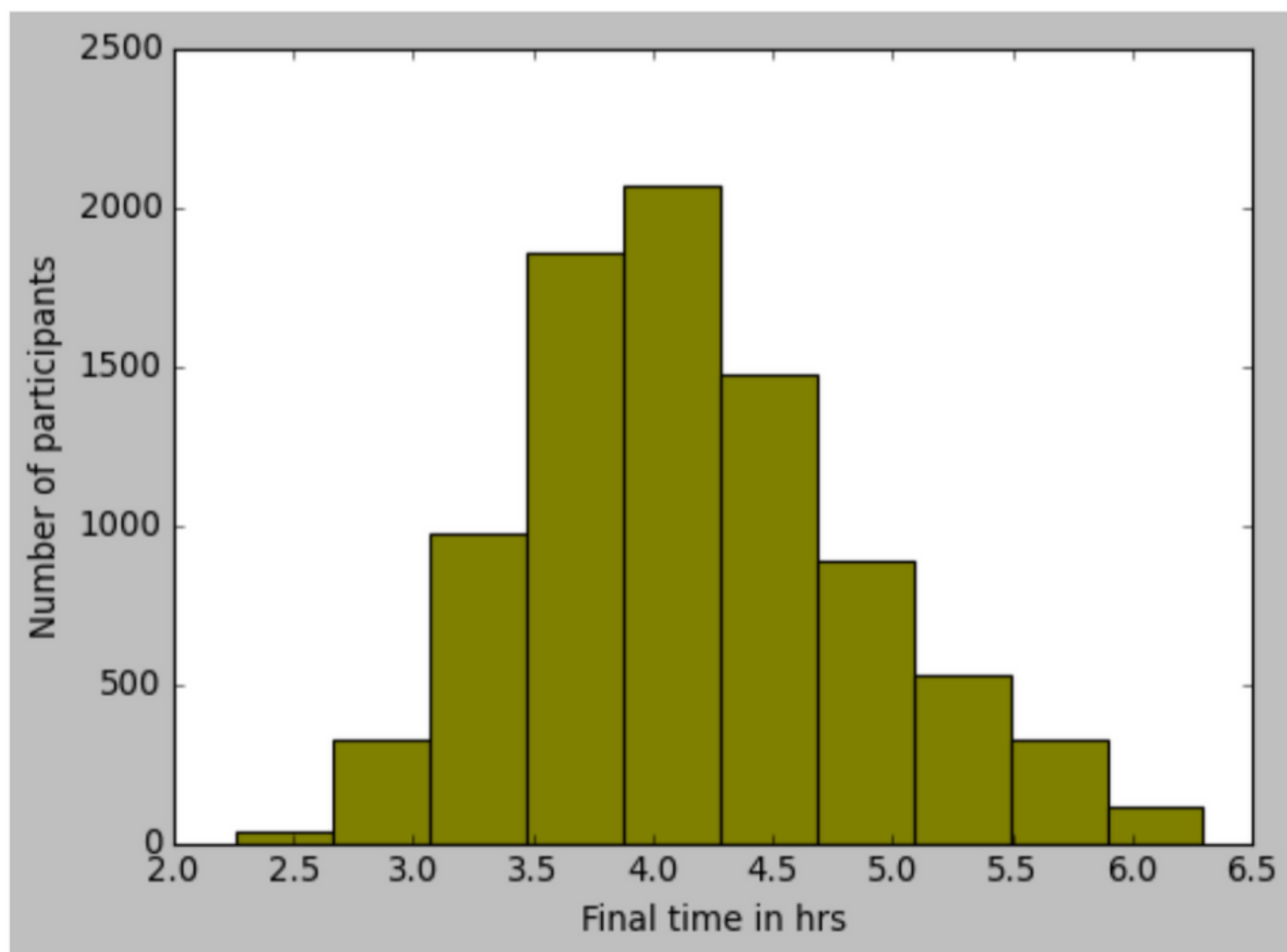
Participants per country colored by age.



Visualicemos la distribución de los tiempos finales de los participantes...

In [13]:

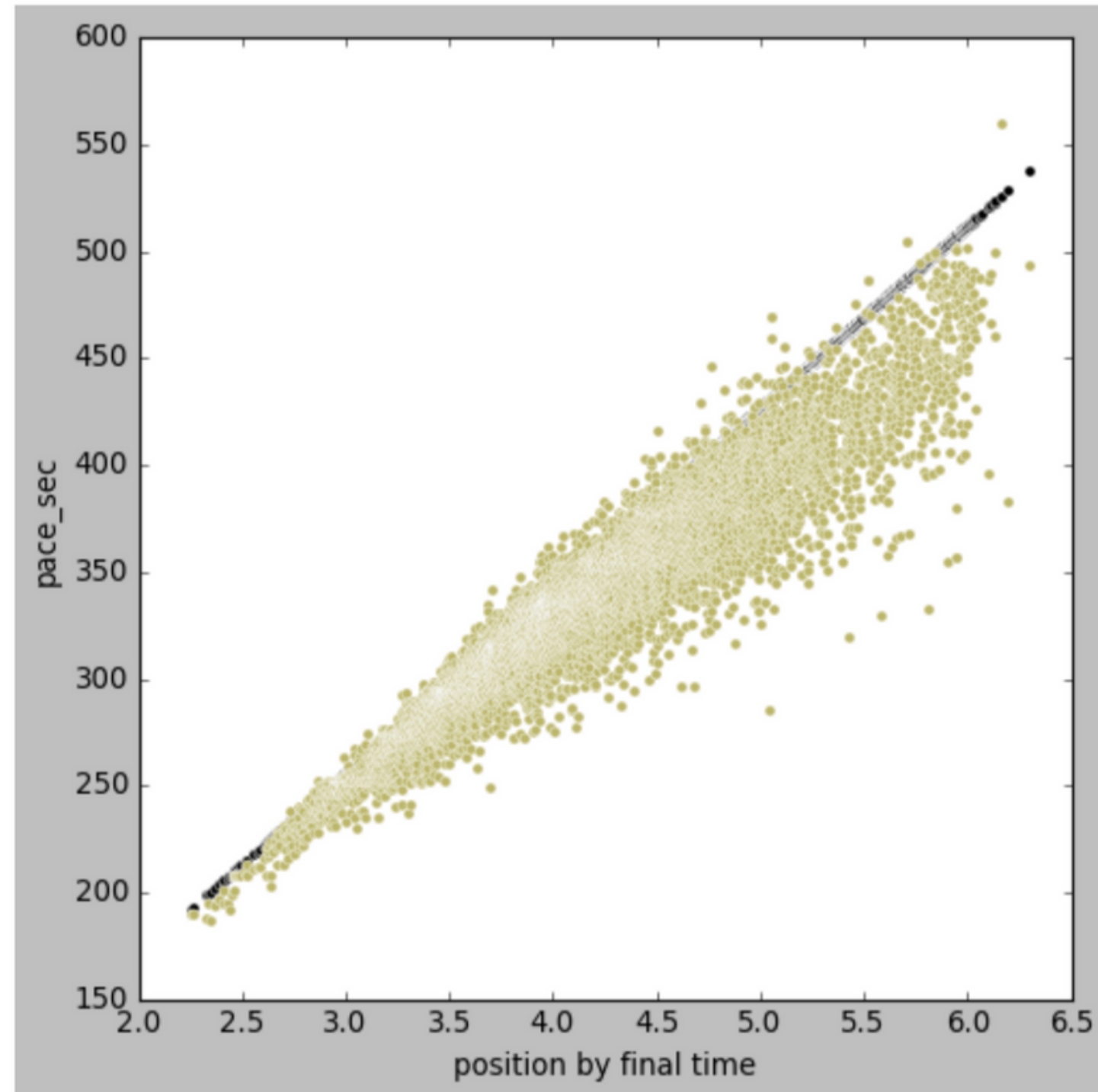
```
plt.figure(figsize=(7,5))
plt.hist(mscw['time_hours'],color='olive')
plt.ylabel('Number of participants')
plt.xlabel('Final time in hrs')
plt.show()
```



Veamos que tal está el ritmo de los participantes:

In [16]:

```
plt.figure(figsize=(7,7))
sns.scatterplot(data=mscw, x="time_hours", y="pace_sec",color='black').set(xlabel='position by final time')
sns.scatterplot(data=mscw, x="time_hours", y="half_pace_sec", color='darkkhaki')#.set(ylim=(0, 6.5))
plt.show()
```



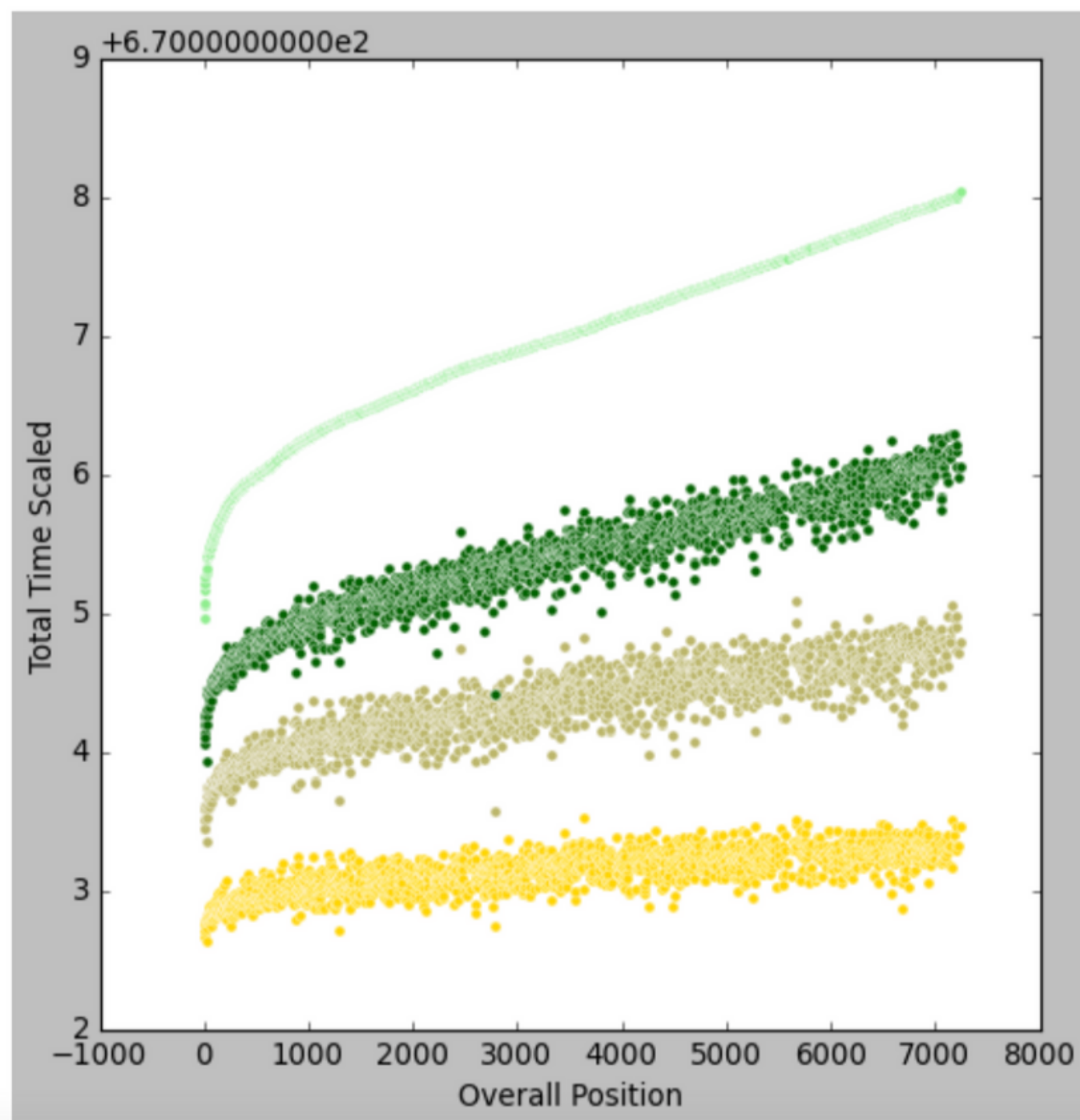
Veamos datos de otro maratón

Aunque los siguientes datos no pertenecen al maratón de Moscú del 2018, el comportamiento observado es replicable en otros maratones.

HONG KONG 

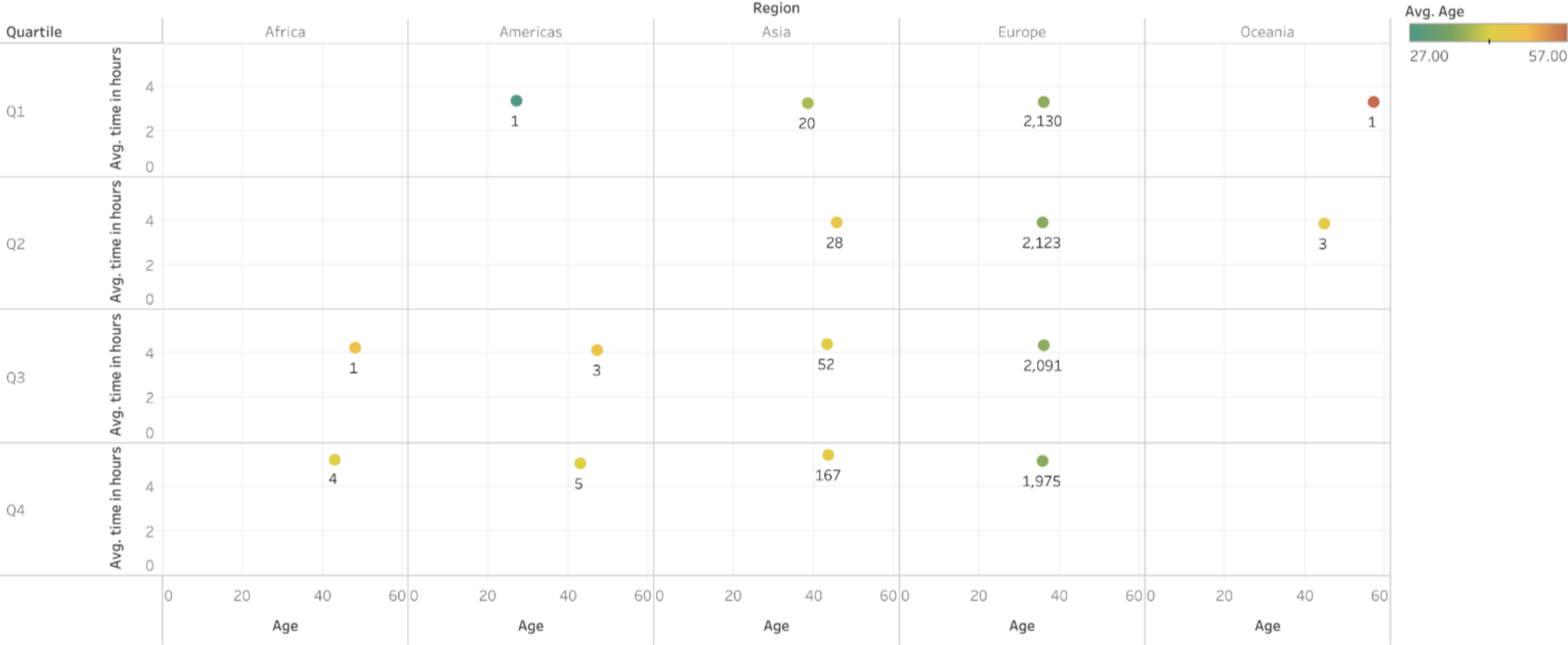
In [18]:

```
plt.figure(figsize=(7,7))
sns.scatterplot(data=china, x="Overall Position", y="10ktimes", color='gold').set(ylabel='Total Time Scaled')
sns.scatterplot(data=china, x="Overall Position", y="halftimes", color='darkkhaki')#.set(ylim=(0, 6.5))
sns.scatterplot(data=china, x="Overall Position", y="30ktimes", color='darkgreen')#.set(ylim=(0, 6.5))
sns.scatterplot(data=china, x="Overall Position", y="finaltimes", color='lightgreen')#.set(ylim=(0, 6.5))
plt.show()
```

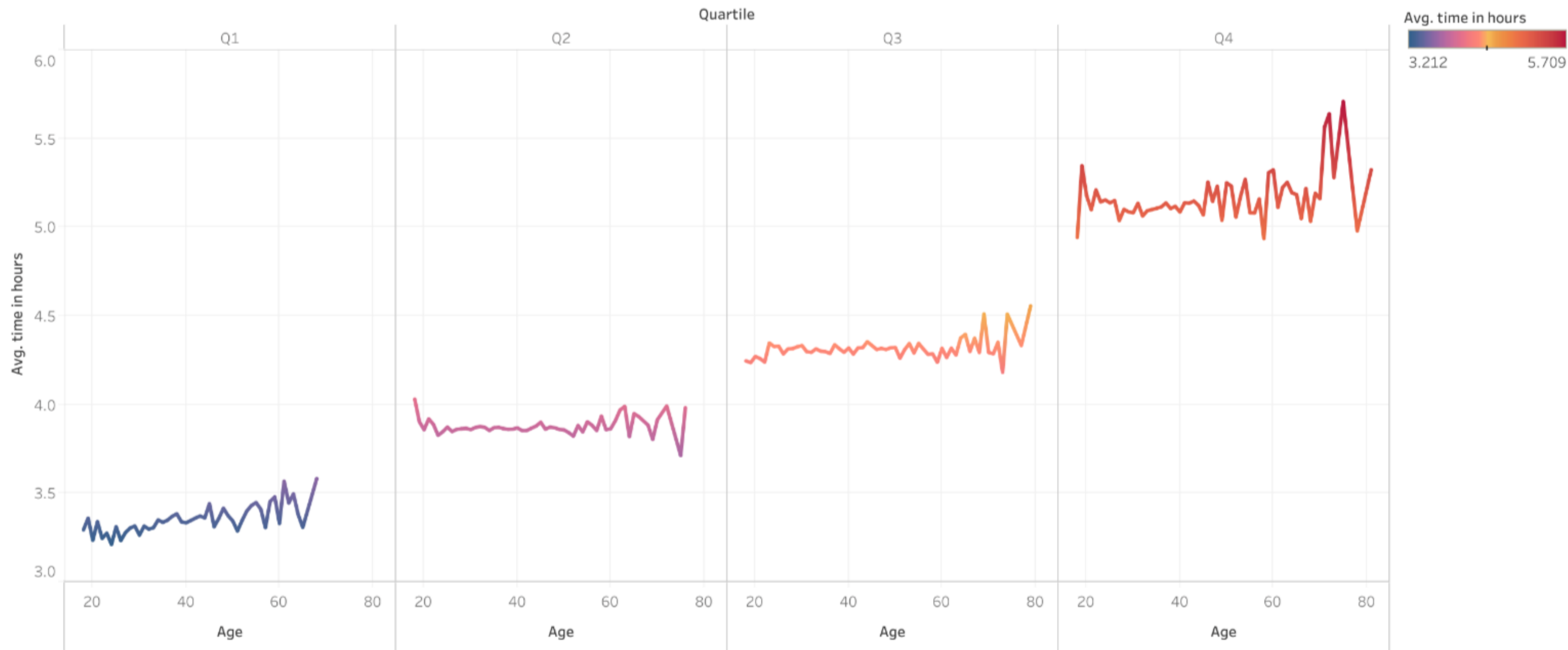


Back to the Kremlin... 

Performance by region and quartile in hours and age, colored by age.



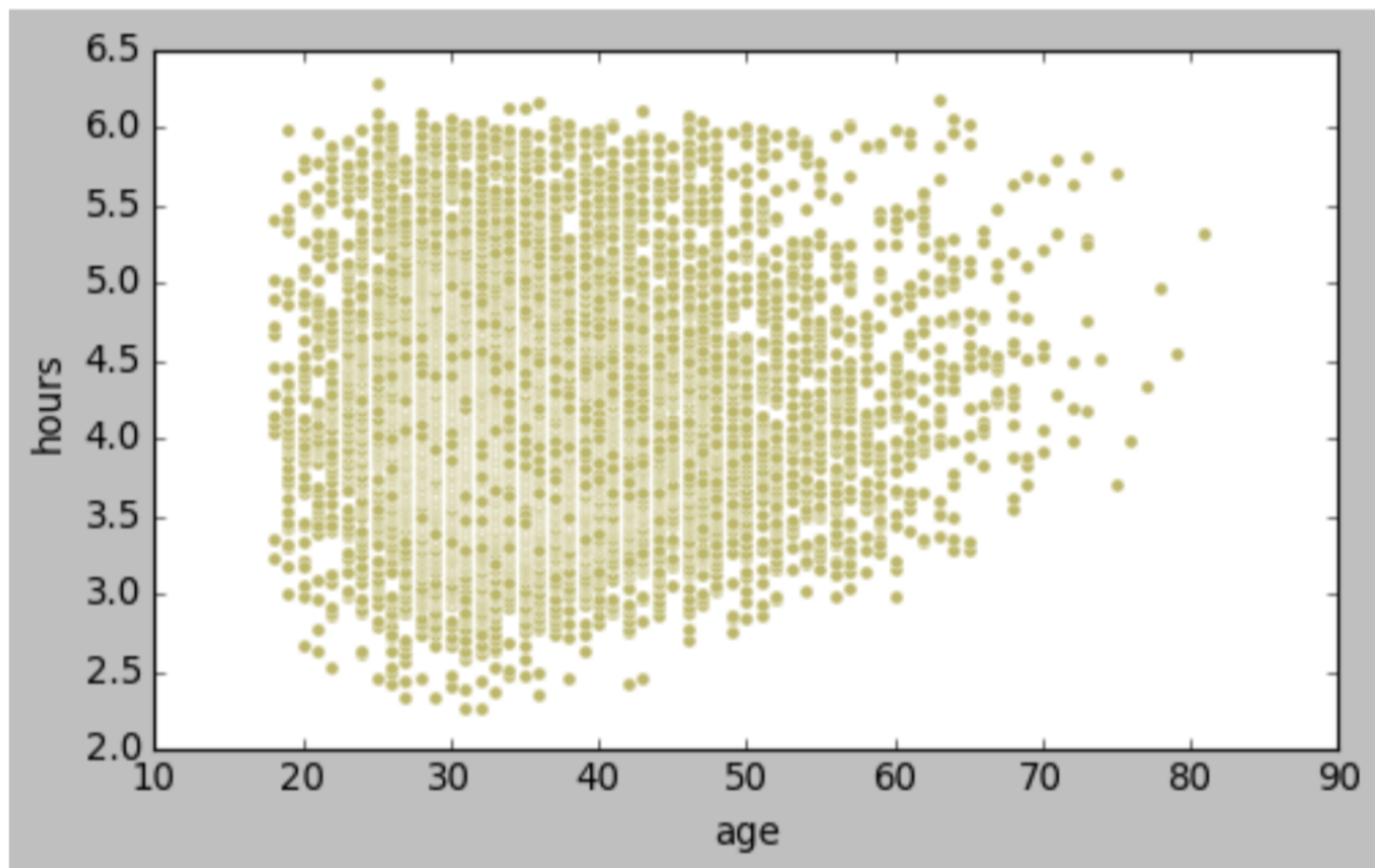
performance by quartile and age colored by pace in seconds



¿Es la edad realmente un factor decisivo?

In [19]:

```
plt.figure(figsize=(7,4))  
sns.scatterplot(data=codes_mscw, x="age", y="time_hours",color='darkkhaki').set( ylabel='hours')  
plt.show()
```



In [22]:

```
age=pd.concat((mscw,bins),axis=1)
pd.pivot_table(age,index='age_range',values= (('time_hours','age')),aggfunc={min, max, np.mean,np.std})
```

Out[22]:

	age				time_hours			
	max	mean	min	std	max	mean	min	std
age_range								
(0-28)	28	25.282163	18	2.530645	6.297778	4.204095	2.331667	0.723213
(29-39)	39	33.579362	29	3.035358	6.160556	4.110546	2.256111	0.690183
(40-49)	49	43.724760	40	2.778067	6.111111	4.173286	2.427222	0.704297
(50-60)	60	53.956882	50	3.072999	6.027222	4.220260	2.843056	0.726409
(61-70)	70	64.360759	61	2.657086	6.191944	4.521145	3.285556	0.680354
(70+)	81	74.050000	71	2.837252	5.813611	4.808153	3.713056	0.671063

Buscamos que la media de los grupos sea significativamente diferente de la media en general. 📈

Recuerda que la media original es: **4.157781**

In [24]:

```
ztest(mscw['time_hours'], value=4.204095)
```

Out[24]:

```
(-6.093653685152342, 1.103622802917605e-09)
```

In [25]:

```
ztest(mscw['time_hours'], value=4.110546)
```

Out[25]:

```
(6.214918750393221, 5.135121389361988e-10)
```

In [26]:

```
ztest(mscw['time_hours'], value=4.173286)
```

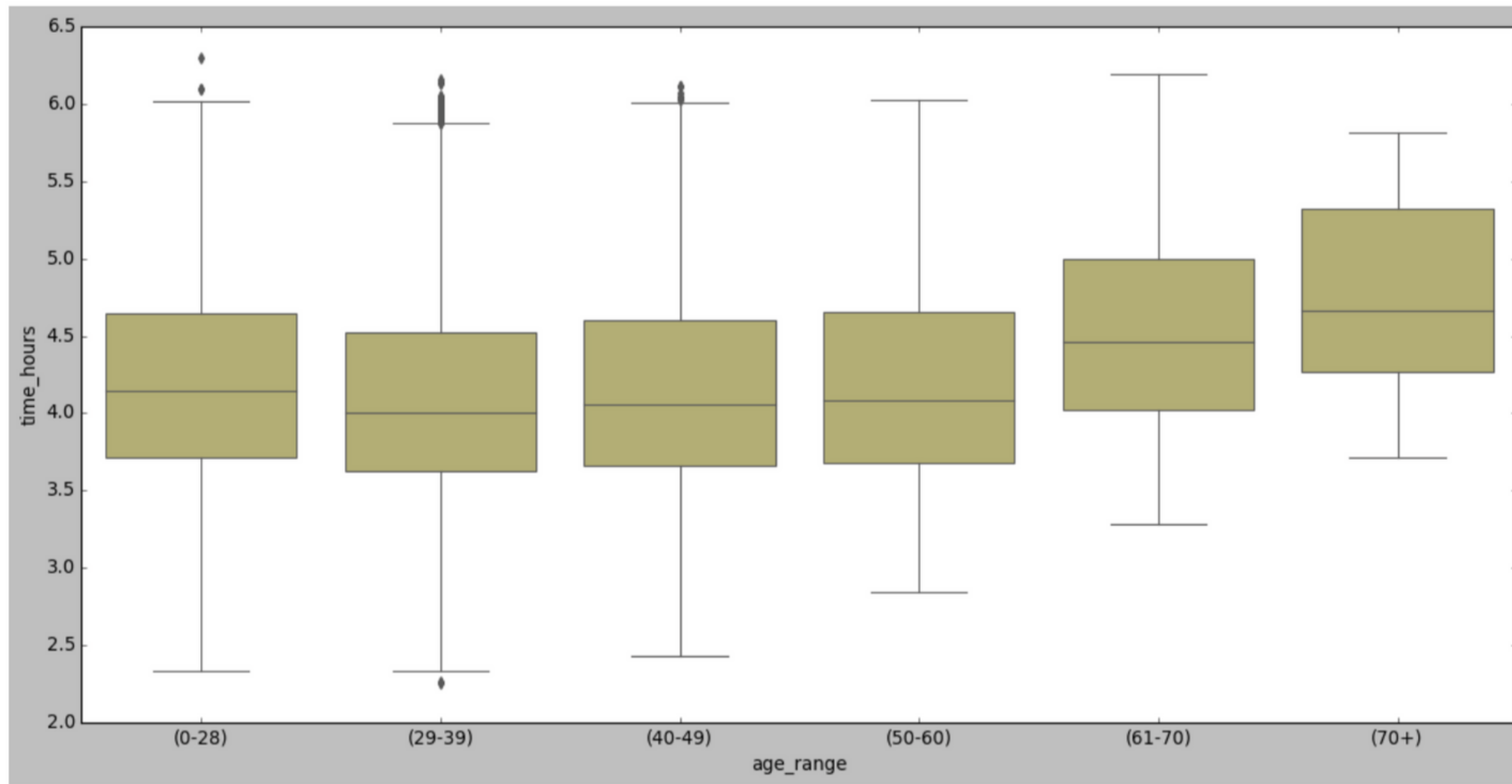
Out[26]:

```
(-2.0400047079668973, 0.041349856820062786)
```

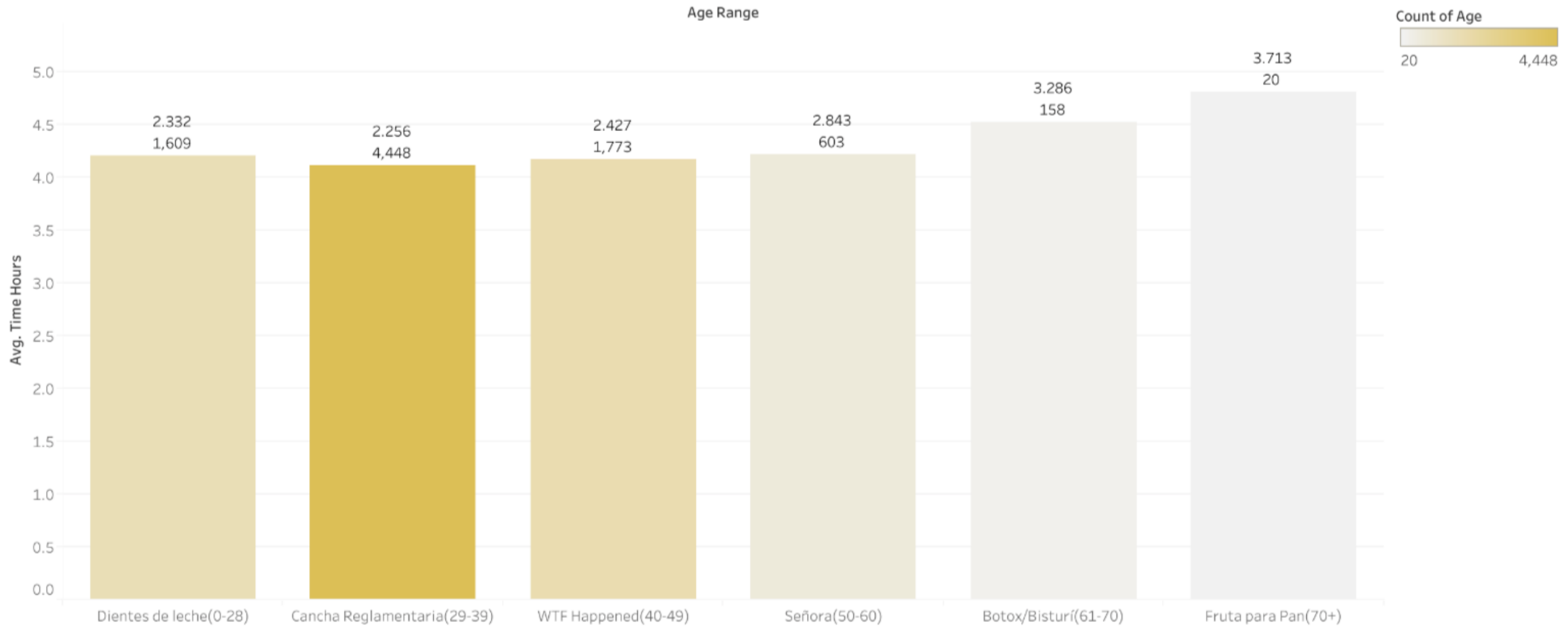
Si, todas son significativamente diferente pero no hay un sentido claro (positivo o negativo en los Z scores) ¿Entonces qué?

In [30]:

```
plt.figure(figsize=(17,8))
sns.boxplot(data=age, x="age_range", y="time_hours",color='darkkhaki')#.set(ylim=(0, 6.5), ylabel='Hours')
plt.show()
```



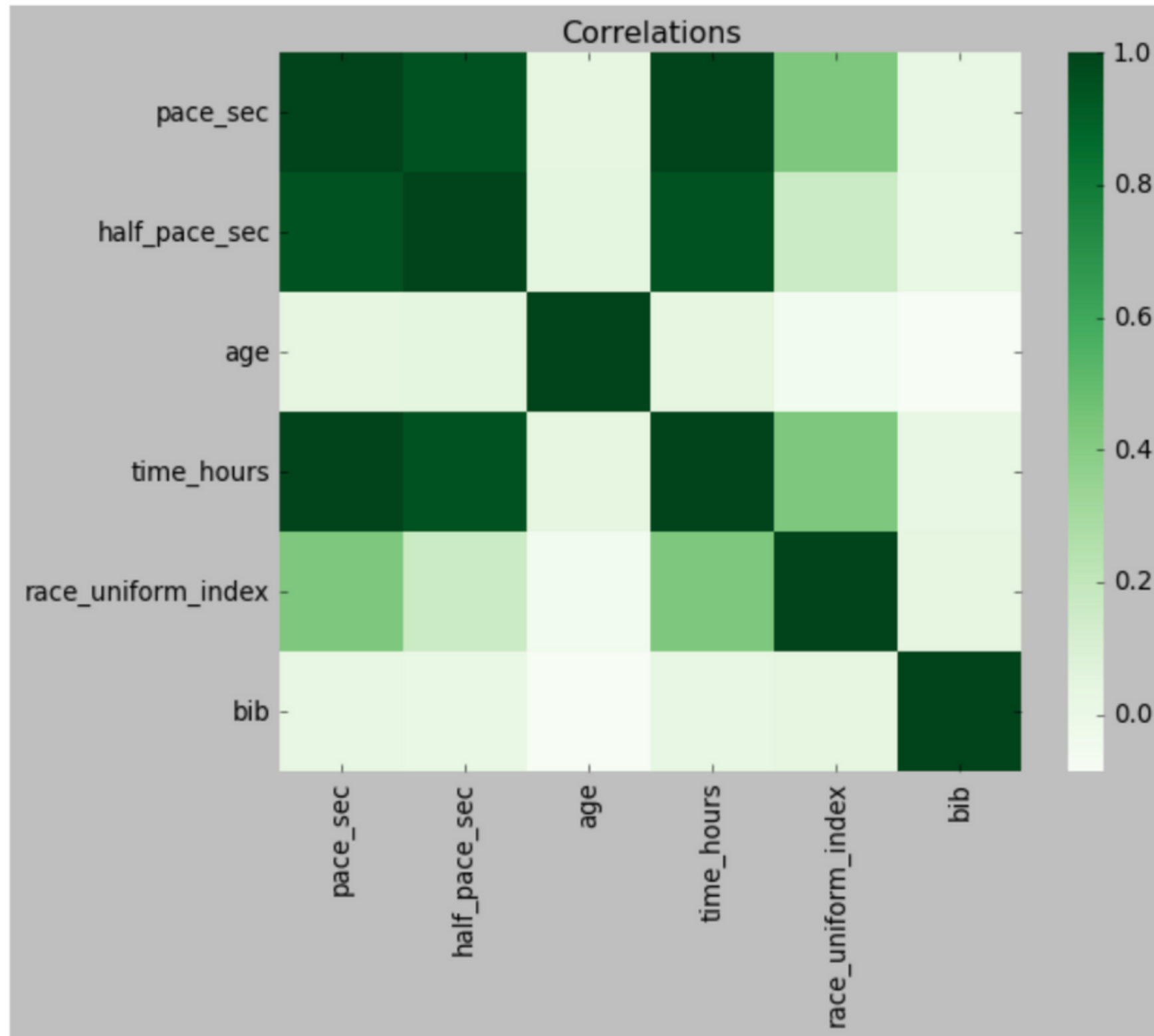
age ranges data



Con este sample no podemos determinar que la edad tenga mucho que ver con el éxito en un maratón.

In [61]:

```
sns.heatmap(matrix,cmap="Greens")  
plt.title('Correlations')  
plt.show()
```



RECAP



1. La edad tal vez no sea un gran factor determinante. Pero si lo fuera estoy en el rango preciso. Es ahora o nunca.

2. El ritmo constante parece ser un factor común en los participantes con mayor éxito.

3. Si eres al menos más rápido que la mitad de los participantes, deberías terminarlo en aproximadamente 4hrs y 9 minutos. Basado en mis entrenamientos hasta ahora. Lo más probable es que no sea más rápido que el 50%. 🙄

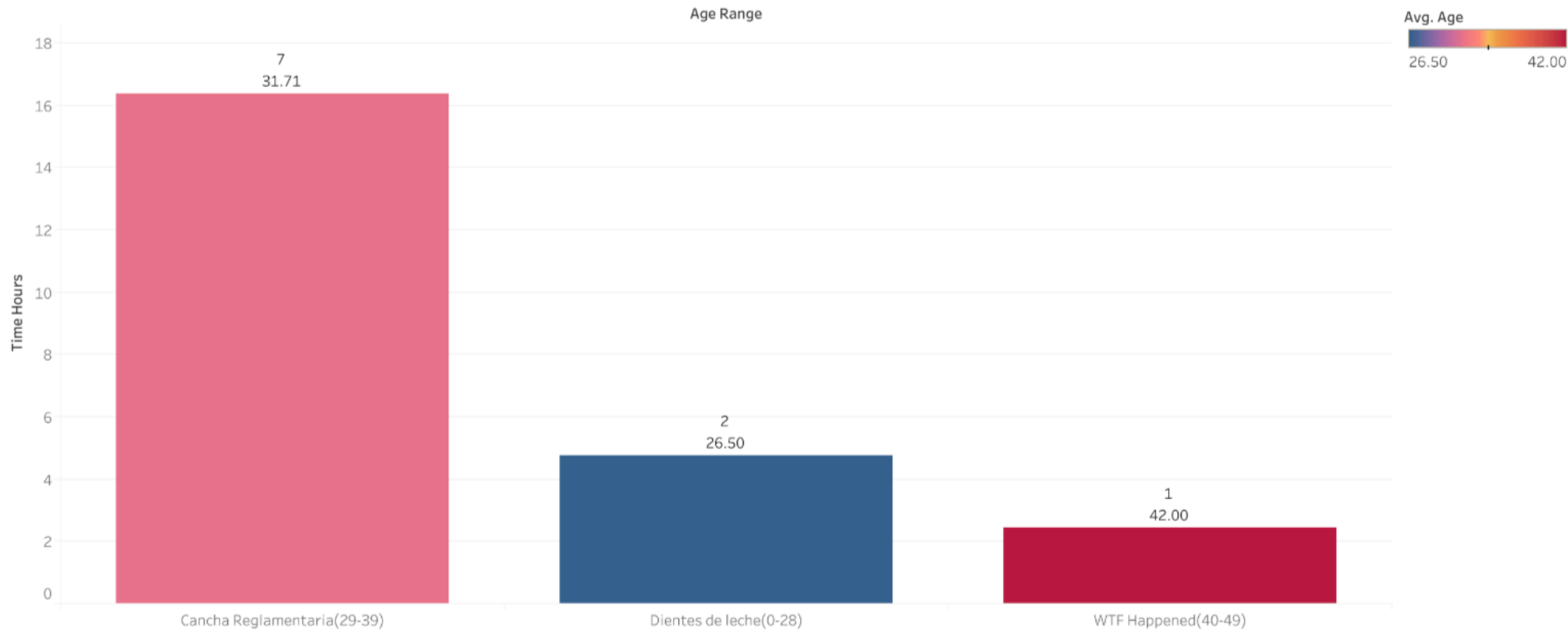
More like the bottom 10%. At least in something I'm a 10-percenter. 😭

4. Los mexicanos literal están en todos lados, osea había 5 mexas en Moscú. 🧑

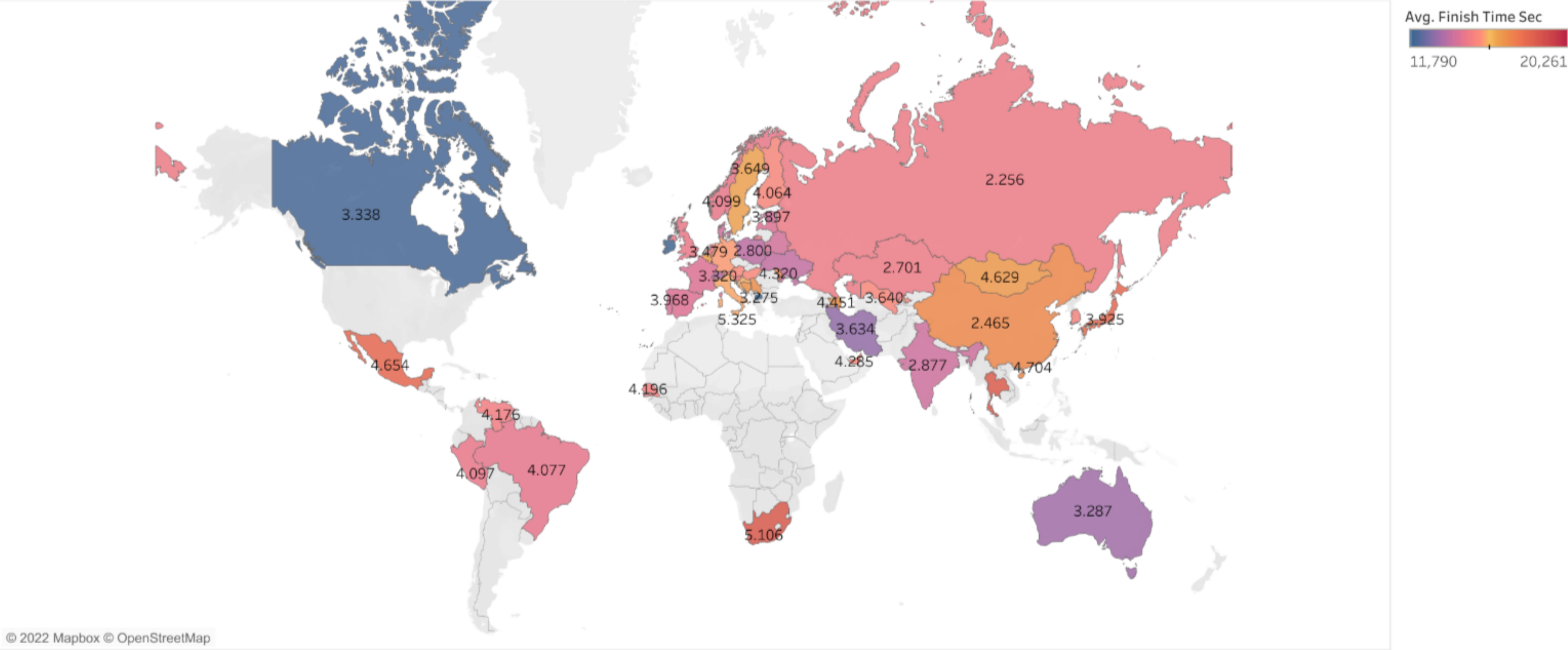


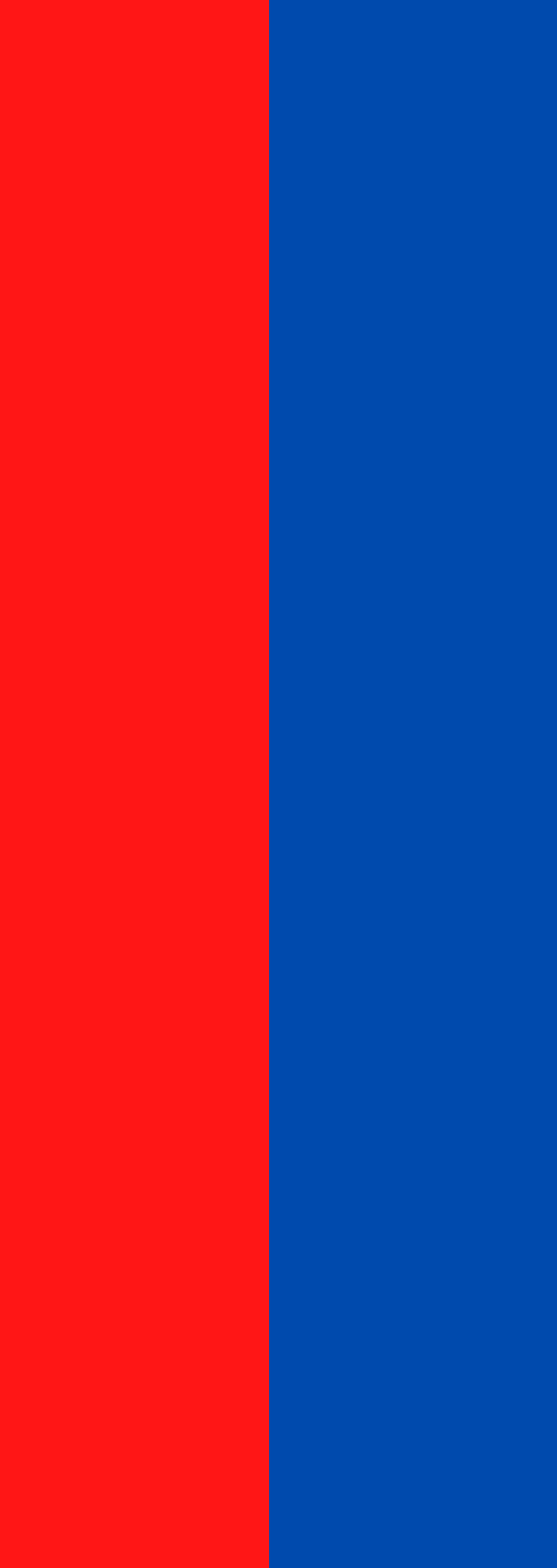
RESULTADOS FINALES

Winners by Age Group and Tally



Average finish time per country labeled by fastest and colored by average.



A decorative graphic on the left side of the slide consisting of two vertical bars: a red bar on the far left and a blue bar immediately to its right.

Спасибо (Gracias!)