Effect of testosterone on blood counts metrics (hgb and hct)

Import data into Pandas ¶

In [2]:

```
import pandas as pd
import numpy as np
```

In [3]:

```
df = pd.read_excel (r'C:\freelancing project\testosterone data collection+age edited.xlsx'
)
```

In [4]:

```
df.head()
```

Out[4]:

	Age in years	Testosterone type	Route of administration	Baseline pre-testosterone hemoglobin/hematocrit	Peak hemoglobin/hematocrit on testosterone	When peak occurred	Management (dose reduction, phlebotomy, etc)	
0	27	Cypionate	IM	14.6/43.4	15.6/45.9	within 72 months of starting	Dose reduced	
1	19	Cypionate	SUBQ	12.7/42.7	13.6/45.8	3mo s/p start date	Dose reduced	a
2	30	Cypionate	IM	12.6/39.8	17.2/50	48 months	Dose reduced	C
3	31	Cypionate	IM	13.3/37.8	16.2/47.5	48 months/p start date	Dose reduced	t
4	26	Cypionate	IM	14.35/42	18/53	23 months	Dose reduced	II

In [5]:
```
df.info ()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234 entries, 0 to 233
Data columns (total 9 columns):
Age in years                                   234 non-null int64
Testosterone type                              234 non-null object
Route of administration                        234 non-null object
Baseline pre-testosterone hemoglobin/hematocrit    234 non-null object
Peak hemoglobin/hematocrit on testosterone     233 non-null object
When peak occurred                             217 non-null object
Management (dose reduction, phlebotomy, etc)   215 non-null object
Comment                                        129 non-null object
Unnamed: 8                                     37 non-null object
dtypes: int64(1), object(8)
memory usage: 16.5+ KB
```

In [ ]:
```

```

Cleaning the data and removing the missing data

Removing missing data and replace None with NaN

In [6]:
```
'DUPLICATE' in df.values
```

Out[6]:
```
False
```

In [7]:
```
'DUPLICATE' in df['Age in years']
'DUPLICATE' in df['Testosterone type']
'DUPLICATE' in df['Route of administration']
'DUPLICATE' in df['Baseline pre-testosterone hemoglobin/hematocrit']
'DUPLICATE' in df['Peak hemoglobin/hematocrit on testosterone']
'DUPLICATE' in df['When peak occurred']
'DUPLICATE' in df['Management (dose reduction, phlebotomy, etc)']
'DUPLICATE' in df['Comment']
```

Out[7]:
```
False
```

Replacing None with NaN

In [8]:
```
df.replace(to_replace =['None', 'none'] ,value =np.nan, inplace=True)
```

Cleaning Testosterone type Column

In [9]:
```
df['Testosterone type'].replace(to_replace =[''] ,value =np.nan, inplace=True)
```

In [10]:

```
df['Testosterone type']
```

Out[10]:

```
0                               Cypionate
1                               Cypionate
2                               Cypionate
3                               Cypionate
4                               Cypionate
5                               Cypionate
6                               Cypionate
7                               Cypionate
8                               Cypionate
9                               Cypionate
10                                      T
11                                      T
12                                      T
13                              Cypionate
14                              Cypionate
15                              Cypionate
16                              Cypionate
17                              Cypionate
18                              Cypionate
19                              Cypionate
20                              Cypionate
21                              Cypionate
22                              Cypionate
23                              Cypionate
24                              Cypionate
25                              Cypionate
26                              Cypionate
27                              Cypionate
28                              Cypionate
29                              Cypionate
                      ...
204                                     T
205                                     T
206                                     T
207                                     T
208                                     T
209                                     T
210                                     T
211                     Testosterone + TC
212                     Testosterone + TC
213                     Testosterone + TC
214                     Testosterone + TC
215                     Testosterone + TC
216                Testosterone enanthate
217                Testosterone enanthate
218                Testosterone enanthate
219                Testosterone enanthate
220                Testosterone enanthate
221                Testosterone enanthate
222                Testosterone enanthate
223                Testosterone enanthate
224                Testosterone enanthate
225                Testosterone enanthate
226    Testosterone enanthate + Cypionate
227    Testosterone enanthate + Cypionate
228                             Cypionate
229                             Cypionate
230                             Cypionate
231                             Cypionate
232                             Cypionate
233                                     T
Name: Testosterone type, Length: 234, dtype: object
```

In [11]:
```
df['Testosterone type'].value_counts()
```

Out[11]:
```
Cypionate                             182
T                                      34
Testosterone enanthate                 10
Testosterone + TC                       5
Testosterone enanthate + Cypionate      2
Cypionate + Testosterone                1
Name: Testosterone type, dtype: int64
```

Cleaning Route of administration column

In [12]:

```
df['Route of administration']
```

```
Out[12]: 0                          IM
         1                        SUBQ
         2                          IM
         3                          IM
         4                          IM
         5                          IM
         6                          IM
         7                          IM
         8                          IM
         9                        SUBQ
         10            Transdermal gel
         11            Transdermal gel
         12            Transdermal gel
         13                         IM
         14                         IM
         15                         IM
         16                         IM
         17                         IM
         18                         IM
         19                         IM
         20                         IM
         21                         IM
         22     IM + Transdermal patch
         23                       SUBQ
         24                       SUBQ
         25                       SUBQ
         26                       SUBQ
         27                         IM
         28                         IM
         29                       SUBQ
                         ...
         204           Transdermal gel
         205         Transdermal patch
         206         Transdermal patch
         207         Transdermal patch
         208         Transdermal patch
         209         Transdermal patch
         210           Transdermal gel
         211      Transdermal gel + IM
         212      Transdermal gel + IM
         213      Transdermal gel + IM
         214      Transdermal gel + IM
         215    Transdermal gel + SUBQ
         216                        IM
         217                        IM
         218                        IM
         219                        IM
         220                        IM
         221                        IM
         222                      SUBQ
         223                      SUBQ
         224                      SUBQ
         225                      SUBQ
         226                        IM
         227                        IM
         228                        IM
         229                        IM
         230                        IM
         231                        IM
         232                        IM
         233           Transdermal gel
         Name: Route of administration, Length: 234, dtype: object
```

```
In [13]: num = df.groupby(['Route of administration']).count()
```

```
In [14]: count = df['Route of administration'].value_counts()
```

In [15]:

```
print(num)
```

```
                         Age in years  Testosterone type  \
Route of administration                                    
IM                                153                153   
IM + SUBQ                           1                  1   
IM + Transdermal patch              4                  4   
Powder/cream                        1                  1   
SUBQ                               37                 37   
Transdermal gel                    28                 28   
Transdermal gel + IM                4                  4   
Transdermal gel + SUBQ              1                  1   
Transdermal patch                   5                  5   

                         Baseline pre-testosterone hemoglobin/hematocrit  \
Route of administration                                                    
IM                                                                   132   
IM + SUBQ                                                              1   
IM + Transdermal patch                                                 3   
Powder/cream                                                           1   
SUBQ                                                                  36   
Transdermal gel                                                       25   
Transdermal gel + IM                                                   3   
Transdermal gel + SUBQ                                                 1   
Transdermal patch                                                      4   

                         Peak hemoglobin/hematocrit on testosterone  \
Route of administration                                               
IM                                                              147   
IM + SUBQ                                                         1   
IM + Transdermal patch                                            3   
Powder/cream                                                      1   
SUBQ                                                             35   
Transdermal gel                                                  23   
Transdermal gel + IM                                              4   
Transdermal gel + SUBQ                                            1   
Transdermal patch                                                 3   

                         When peak occurred  \
Route of administration                       
IM                                      136   
IM + SUBQ                                 1   
IM + Transdermal patch                    3   
Powder/cream                              1   
SUBQ                                     31   
Transdermal gel                          24   
Transdermal gel + IM                      4   
Transdermal gel + SUBQ                    1   
Transdermal patch                         3   

                         Management (dose reduction, phlebotomy, etc)  \
Route of administration                                                 
IM                                                                140   
IM + SUBQ                                                           1   
IM + Transdermal patch                                              3   
Powder/cream                                                        1   
SUBQ                                                               33   
Transdermal gel                                                    27   
Transdermal gel + IM                                                4   
Transdermal gel + SUBQ                                              1   
Transdermal patch                                                   5   

                         Comment  Unnamed: 8  
Route of administration                       
IM                            68          21  
IM + SUBQ                      0           0  
IM + Transdermal patch         3          0
Powder/cream                   1          1
SUBQ                          19          2
Transdermal gel               28          8
Transdermal gel + IM           4          2
Transdermal gel + SUBQ         1          0
Transdermal patch              5          3
```

In [16]:
```
print(count)
```

```
IM                          153
SUBQ                         37
Transdermal gel              28
Transdermal patch             5
IM + Transdermal patch        4
Transdermal gel + IM          4
IM + SUBQ                     1
Transdermal gel + SUBQ        1
Powder/cream                  1
Name: Route of administration, dtype: int64
```

Cleaning Management (dose reduction, phlebotomy, etc) Column

In [ ]:

In [17]:
```
df['Management (dose reduction, phlebotomy, etc)'].count()
```

Out[17]:
```
215
```

Cleaning When peak occurred Column

Removing Months from When peak occurred Column and replacing None with NaN

In [18]:
```
df['When peak occurred'] = df['When peak occurred'].str.replace('months', '')
df['When peak occurred'] = df['When peak occurred'].str.replace('month after', '')
df['When peak occurred'] = df['When peak occurred'].str.replace('after', '')
df['When peak occurred'] = df['When peak occurred'].str.replace('month (days)', '')
df['When peak occurred'] = df['When peak occurred'].str.replace('mo s/p start date', '')
df['When peak occurred'] = df['When peak occurred'].str.replace('s/p start date', '')
df['When peak occurred'] = df['When peak occurred'].str.replace('of starting', '')
df['When peak occurred'] = df['When peak occurred'].str.replace('within', '')

df['When peak occurred'].replace(to_replace =['None', 'none'] ,value =np.nan, inplace=True
)
```

replacing < 1 month (days) with value = 0.5 in When peak occurred Column

In [19]:
```
df['When peak occurred'].replace(to_replace ='< 1 month (days)' ,value = 1)
df['When peak occurred']= pd.to_numeric(df['When peak occurred'], errors='coerce')
df['When peak occurred'].dtypes
```

Out[19]:
```
dtype('float64')
```

Cleaning Baseline pre-testosterone hemoglobin/hematocrit column and Peak hemoglobin/hematocrit on testosterone column

In [20]:

```
#cleaning Baseline pre-testosterone hemoglobin/hematocrit column
#replacing None with NaN and removing words
#convert data type tp float
df['Baseline pre-testosterone hemoglobin/hematocrit'].replace(to_replace =['None', 'none']
,value =np.nan,inplace=True)
df['Baseline pre-testosterone hemoglobin/hematocrit'].replace(to_replace=r'^[a-zA-Z\s,?()]
+', value='', regex=True,inplace=True)
df['Baseline pre-testosterone hemoglobin/hematocrit'] = df['Baseline pre-testosterone hemo
globin/hematocrit'].str.replace('2007 in FL', '')
df['Baseline pre-testosterone hemoglobin/hematocrit'] = df['Baseline pre-testosterone hemo
globin/hematocrit'].str.replace('2013', '')
df['Baseline pre-testosterone hemoglobin/hematocrit'] = df['Baseline pre-testosterone hemo
globin/hematocrit'].str.replace('2015', '')
df['Baseline pre-testosterone hemoglobin/hematocrit'].replace(to_replace =['None', 'none',
''] ,value =np.nan,inplace=True)
df['Baseline pre hgb/htc'] = df['Baseline pre-testosterone hemoglobin/hematocrit']
```

In [ ]:

```

```

In [21]:

```
df['Baseline pre-testosterone hemoglobin/hematocrit'] = df['Baseline pre-testosterone hemo
globin/hematocrit'].apply(lambda x: np.nan if x is np.nan else eval(x))
```

In [22]:

```
df['Baseline pre-testosterone hemoglobin/hematocrit']
```

Out[22]:

```
0      0.336406
1      0.297424
2      0.316583
3      0.351852
4      0.341667
5      0.328016
6      0.339168
7      0.320652
8           NaN
9      0.339726
10     0.331620
11     0.337709
12     0.322957
13     0.331742
14          NaN
15          NaN
16     0.322816
17     0.341146
18     0.355316
19     0.340686
20     0.328321
21     0.322654
22     0.340979
23     0.346934
24     0.322799
25          NaN
26     0.328283
27     0.338971
28          NaN
29          NaN
         ...   
204    0.342394
205    0.342371
206    0.297229
207    0.330657
208         NaN
209    0.339332
210         NaN
211    0.340450
212    0.332606
213    0.333333
214         NaN
215    0.326260
216         NaN
217    0.331828
218    0.337070
219    0.341608
220    0.334399
221    0.332772
222    0.341204
223    0.328228
224    0.347123
225    0.333333
226    0.329582
227         NaN
228         NaN
229         NaN
230         NaN
231         NaN
232         NaN
233    0.337696
Name: Baseline pre-testosterone hemoglobin/hematocrit, Length: 234, dtype: float64
```

In [23]:

```
#cleaning Peak hemoglobin/hematocrit on testosterone column
#replacing None with NaN and removing words
#convert data type tp float

df['Peak hemoglobin/hematocrit on testosterone'].replace(to_replace =['None', 'none'] ,value =np.nan,inplace=True)
df['Peak hemoglobin/hematocrit on testosterone'].replace(to_replace=r'^[a-zA-Z\s,?()]+', value='', regex=True,inplace=True)
df['Peak hemoglobin/hematocrit on testosterone'].replace(to_replace =['None', 'none', ''] ,value =np.nan,inplace=True)
df['Post hgb/htc']= df['Peak hemoglobin/hematocrit on testosterone']
```

In [24]:

```
df['Peak hemoglobin/hematocrit on testosterone'] = df['Peak hemoglobin/hematocrit on testosterone'].apply(lambda x: np.nan if x is np.nan else eval(x))
```

```
In [25]: df['Peak hemoglobin/hematocrit on testosterone']
```

```
Out[25]: 0      0.339869
         1      0.296943
         2      0.344000
         3      0.341053
         4      0.339623
         5      0.323413
         6      0.345924
         7      0.339286
         8      0.344538
         9      0.344902
         10     0.335260
         11     0.340000
         12     0.338000
         13          NaN
         14          NaN
         15          NaN
         16          NaN
         17          NaN
         18          NaN
         19          NaN
         20          NaN
         21          NaN
         22          NaN
         23          NaN
         24          NaN
         25          NaN
         26          NaN
         27     0.327519
         28     0.324803
         29     0.343750
                  ...
         204    0.333333
         205         NaN
         206         NaN
         207    0.336557
         208    0.327309
         209    0.341404
         210    0.336493
         211    0.337900
         212    0.323789
         213    0.337838
         214    0.335878
         215    0.324211
         216    0.323251
         217    0.325142
         218    0.340726
         219    0.328094
         220    0.327751
         221         NaN
         222    0.342742
         223    0.343750
         224    0.344907
         225         NaN
         226    0.330073
         227    0.319202
         228    0.345009
         229         NaN
         230         NaN
         231         NaN
         232    0.345572
         233    0.338824
         Name: Peak hemoglobin/hematocrit on testosterone, Length: 234, dtype: float64
```

Exploratory Data Analysis

Mean age

-Median time to peak

-Mean hgb/hct pre and post

-Pecent over 17.6 hgb and 50% hct

-Mean change in hemoglobin and hematocrit

-Percent with each type of testosterone (IM, oral, transdermal)

-Mean peak with each type

-% with each management type (none, lower dose or phlebotomy)

Mean age

```
In [26]: # skip the Nan values while finding the median & mean
         df['Age in years'].mean()
```

```
Out[26]: 29.145299145299145
```

Median time to peak

```
In [27]: df['When peak occurred'].median(axis = 0, skipna = True)
```

```
Out[27]: 21.0
```

Mean hgb/hct pre and post

```
In [28]: #Mean hgb/hct pre
         df['Baseline pre-testosterone hemoglobin/hematocrit'].mean()
```

```
Out[28]: 0.3342346048156397
```

```
In [29]: #mean hgb/hct post
         df['Peak hemoglobin/hematocrit on testosterone'].mean()
```

```
Out[29]: 0.33351869480645685
```

Pecent over 17.6 hgb and 50% hct

```
In [30]: (df['Peak hemoglobin/hematocrit on testosterone'] > 0.352 ).mean()*100
```

```
Out[30]: 3.8461538461538463
```

In [31]:

```
df[df['Peak hemoglobin/hematocrit on testosterone'] > 0.352]
```

Out[31]:

	Age in years	Testosterone type	Route of administration	Baseline pre-testosterone hemoglobin/hematocrit	Peak hemoglobin/hematocrit on testosterone	When peak occurred	Management (dose reduction, phlebotomy, etc)
31	56	Cypionate	IM	0.344550	0.352490	60.0	NONE
55	29	Cypionate	IM	0.345013	0.353881	12.0	NONE
71	39	Cypionate	IM	0.354067	0.353430	18.0	NONE
99	25	Cypionate	IM	0.363218	0.352818	36.0	NONE
108	32	Cypionate	IM	0.338346	0.357918	60.0	NONE
110	56	Cypionate	IM	NaN	0.352113	12.0	NONE
124	26	Cypionate	IM	NaN	0.356481	22.0	NONE
132	29	Cypionate	IM	0.348469	0.356195	21.0	NONE
203	32	T	Transdermal gel	0.348581	0.353516	32.0	NONE

Mean change in hemoglobin and hematocrit

In [32]:

```
# new data frame with split value columns
new = df["Baseline pre hgb/htc"].str.split("/", n = 1, expand = True)

# making separate first pre hgb column from new data frame
df["Pre Hgb"]= new[0]
df["Pre Hgb"] = pd.to_numeric(df["Pre Hgb"], downcast="float")

# making separate last pre htc column from new data frame
df["Pre Htc"]= new[1]
df["Pre Htc"] = pd.to_numeric(df["Pre Htc"], downcast="float")
```

In [33]:

```
df["Pre Hgb"].mean()
```

Out[33]:

```
13.473546028137207
```

In [34]:

```
df["Pre Htc"].mean()
```

Out[34]:

```
40.30673599243164
```

In [35]:

```

# new data frame with split value columns
new1 = df["Post hgb/htc"].str.split("/", n = 1, expand = True)

# making separate first Post hgb column from new data frame
df["Post Hgb"]= new1[0]
df["Post Hgb"] = pd.to_numeric(df["Post Hgb"], downcast="float")


# making separate last Post htc column from new data frame
df["Post Htc"]= new1[1]
df["Post Htc"] = pd.to_numeric(df["Post Htc"], downcast="float")
```

In [36]:

```
df["Post Hgb"].mean()
```

Out[36]:

```
15.74777889251709
```

In [37]:

```
df["Post Htc"].mean()
```

Out[37]:

```
47.206905364990234
```

In [38]:

```
df
```

Out[38]:

	Age in years	Testosterone type	Route of administration	Baseline pre-testosterone hemoglobin/hematocrit	Peak hemoglobin/hematocrit on testosterone	When peak occurred	Management (dose reduction, phlebotomy, etc)
0	27	Cypionate	IM	0.336406	0.339869	72.0	Dose reduced
1	19	Cypionate	SUBQ	0.297424	0.296943	3.0	Dose reduced
2	30	Cypionate	IM	0.316583	0.344000	48.0	Dose reduced
3	31	Cypionate	IM	0.351852	0.341053	NaN	Dose reduced
4	26	Cypionate	IM	0.341667	0.339623	23.0	Dose reduced
5	33	Cypionate	IM	0.328016	0.323413	19.0	Dose reduced
6	31	Cypionate	IM	0.339168	0.345924	63.0	Dose reduced
7	27	Cypionate	IM	0.320652	0.339286	20.0	Dose reduced
8	37	Cypionate	IM	NaN	0.344538	192.0	Dose reduced
9	28	Cypionate	SUBQ	0.339726	0.344902	9.0	Dose reduced
10	43	T	Transdermal gel	0.331620	0.335260	24.0	Dose reduced
11	36	T	Transdermal gel	0.337709	0.340000	24.0	Dose reduced
12	28	T	Transdermal gel	0.322957	0.338000	9.0	Dose reduced
13	18	Cypionate	IM	0.331742	NaN	NaN	NaN
14	23	Cypionate	IM	NaN	NaN	NaN	NaN
15	31	Cypionate	IM	NaN	NaN	NaN	NaN

	Age in years	Testosterone type	Route of administration	Baseline pre-testosterone hemoglobin/hematocrit	Peak hemoglobin/hematocrit on testosterone	When peak occurred	Management (dose reduction, phlebotomy, etc)
16	31	Cypionate	IM	0.322816	NaN	NaN	NaN
17	19	Cypionate	IM	0.341146	NaN	NaN	NaN
18	19	Cypionate	IM	0.355316	NaN	NaN	NaN
19	23	Cypionate	IM	0.340686	NaN	NaN	NaN
20	18	Cypionate	IM	0.328321	NaN	NaN	NaN
21	32	Cypionate	IM	0.322654	NaN	NaN	NaN
22	18	Cypionate	IM + Transdermal patch	0.340979	NaN	NaN	NaN
23	20	Cypionate	SUBQ	0.346934	NaN	NaN	NaN
24	23	Cypionate	SUBQ	0.322799	NaN	NaN	NaN
25	26	Cypionate	SUBQ	NaN	NaN	NaN	NaN
26	18	Cypionate	SUBQ	0.328283	NaN	NaN	NaN
27	21	Cypionate	IM	0.338971	0.327519	32.0	NONE
28	33	Cypionate	IM	NaN	0.324803	30.0	NONE
29	35	Cypionate	SUBQ	NaN	0.343750	36.0	NONE
...	...	...	...	...	...	...	...
204	25	T	Transdermal gel	0.342394	0.333333	10.0	NONE
205	18	T	Transdermal patch	0.342371	NaN	NaN	NONE
206	18	T	Transdermal patch	0.297229	NaN	NaN	NONE
207	47	T	Transdermal patch	0.330657	0.336557	12.0	NONE
208	20	T	Transdermal patch	NaN	0.327309	14.0	NONE
209	23	T	Transdermal patch	0.339332	0.341404	18.0	NONE

	Age in years	Testosterone type	Route of administration	Baseline pre-testosterone hemoglobin/hematocrit	Peak hemoglobin/hematocrit on testosterone	When peak occurred	Management (dose reduction, phlebotomy, etc)
210	26	T	Transdermal gel	NaN	0.336493	36.0	NONE
211	54	Testosterone + TC	Transdermal gel + IM	0.340450	0.337900	48.0	NONE
212	24	Testosterone + TC	Transdermal gel + IM	0.332606	0.323789	18.0	NONE
213	27	Testosterone + TC	Transdermal gel + IM	0.333333	0.337838	33.0	NONE
214	22	Testosterone + TC	Transdermal gel + IM	NaN	0.335878	4.0	NONE
215	22	Testosterone + TC	Transdermal gel + SUBQ	0.326260	0.324211	18.0	NONE
216	70	Testosterone enanthate	IM	NaN	0.323251	19.0	NONE
217	39	Testosterone enanthate	IM	0.331828	0.325142	132.0	NONE
218	27	Testosterone enanthate	IM	0.337070	0.340726	19.0	NONE
219	31	Testosterone enanthate	IM	0.341608	0.328094	24.0	NONE
220	36	Testosterone enanthate	IM	0.334399	0.327751	43.0	NONE
221	26	Testosterone enanthate	IM	0.332772	NaN	NaN	NONE
222	39	Testosterone enanthate	SUBQ	0.341204	0.342742	35.0	NONE
223	24	Testosterone enanthate	SUBQ	0.328228	0.343750	4.0	NONE

	Age in years	Testosterone type	Route of administration	Baseline pre-testosterone hemoglobin/hematocrit	Peak hemoglobin/hematocrit on testosterone	When peak occurred	Management (dose reduction, phlebotomy, etc)
224	18	Testosterone enanthate	SUBQ	0.347123	0.344907	15.0	NONE
225	26	Testosterone enanthate	SUBQ	0.333333	NaN	NaN	NONE
226	40	Testosterone enanthate + Cypionate	IM	0.329582	0.330073	17.0	NONE
227	22	Testosterone enanthate + Cypionate	IM	NaN	0.319202	13.0	NONE
228	54	Cypionate	IM	NaN	0.345009	60.0	NONE
229	30	Cypionate	IM	NaN	NaN	NaN	NaN
230	24	Cypionate	IM	NaN	NaN	NaN	NaN
231	26	Cypionate	IM	NaN	NaN	NaN	NaN
232	27	Cypionate	IM	NaN	0.345572	58.0	NaN
233	22	T	Transdermal gel	0.337696	0.338824	14.0	NaN

234 rows × 15 columns

Percent with each type of testosterone (IM, oral, transdermal)

In [39]:

```
df['no_missing_column'] = 'non-missing'
dfg = df.groupby(['Route of administration', 'Testosterone type'])['no_missing_column'].co
unt()
dfg.groupby(level=0).apply(lambda x:100 * x / float(x.sum()))
```

Out[39]:

```
Route of administration  Testosterone type
IM                       Cypionate                              94.771242
                         Testosterone enanthate                  3.921569
                         Testosterone enanthate + Cypionate      1.307190
IM + SUBQ                Cypionate                             100.000000
IM + Transdermal patch   Cypionate                              75.000000
                         Cypionate + Testosterone               25.000000
Powder/cream             T                                     100.000000
SUBQ                     Cypionate                              89.189189
                         Testosterone enanthate                 10.810811
Transdermal gel          T                                     100.000000
Transdermal gel + IM     Testosterone + TC                     100.000000
Transdermal gel + SUBQ   Testosterone + TC                     100.000000
Transdermal patch        T                                     100.000000
Name: no_missing_column, dtype: float64
```

In [40]:

```
df['no_missing_column'] = 'non-missing'
dfg = df.groupby(['Route of administration', 'Testosterone type'])['no_missing_column'].co
unt()/234 *100
dfg
```

Out[40]:

```
Route of administration  Testosterone type
IM                       Cypionate                             61.965812
                         Testosterone enanthate                 2.564103
                         Testosterone enanthate + Cypionate     0.854701
IM + SUBQ                Cypionate                              0.427350
IM + Transdermal patch   Cypionate                              1.282051
                         Cypionate + Testosterone               0.427350
Powder/cream             T                                      0.427350
SUBQ                     Cypionate                             14.102564
                         Testosterone enanthate                 1.709402
Transdermal gel          T                                     11.965812
Transdermal gel + IM     Testosterone + TC                      1.709402
Transdermal gel + SUBQ   Testosterone + TC                      0.427350
Transdermal patch        T                                      2.136752
Name: no_missing_column, dtype: float64
```

In [41]:

```
df['Testosterone type'].value_counts()
```

Out[41]:

```
Cypionate                             182
T                                      34
Testosterone enanthate                 10
Testosterone + TC                       5
Testosterone enanthate + Cypionate      2
Cypionate + Testosterone                1
Name: Testosterone type, dtype: int64
```

In [42]:

```
df['Testosterone type'].value_counts()/234 *100
```

Out[42]:

```
Cypionate                             77.777778
T                                     14.529915
Testosterone enanthate                 4.273504
Testosterone + TC                      2.136752
Testosterone enanthate + Cypionate     0.854701
Cypionate + Testosterone               0.427350
Name: Testosterone type, dtype: float64
```

```
In [43]: df['Route of administration'].value_counts()
```

```
Out[43]: IM                         153
         SUBQ                        37
         Transdermal gel             28
         Transdermal patch            5
         IM + Transdermal patch       4
         Transdermal gel + IM         4
         IM + SUBQ                    1
         Transdermal gel + SUBQ       1
         Powder/cream                 1
         Name: Route of administration, dtype: int64
```

```
In [44]: df['Route of administration'].value_counts()/234 *100
```

```
Out[44]: IM                         65.384615
         SUBQ                       15.811966
         Transdermal gel            11.965812
         Transdermal patch           2.136752
         IM + Transdermal patch      1.709402
         Transdermal gel + IM        1.709402
         IM + SUBQ                   0.427350
         Transdermal gel + SUBQ      0.427350
         Powder/cream                0.427350
         Name: Route of administration, dtype: float64
```

Mean peak with each type of testosterone (IM, oral, transdermal)

```
In [45]: dfg = df.groupby(['Route of administration', 'Testosterone type'])['Peak hemoglobin/hemato
         crit on testosterone'].mean()
         dfg
```

```
Out[45]: Route of administration  Testosterone type
         IM                       Cypionate                                 0.333756
                                  Testosterone enanthate                    0.328993
                                  Testosterone enanthate + Cypionate        0.324638
         IM + SUBQ                Cypionate                                 0.328160
         IM + Transdermal patch   Cypionate                                 0.327459
                                  Cypionate + Testosterone                  0.344086
         Powder/cream             T                                         0.330918
         SUBQ                     Cypionate                                 0.332593
                                  Testosterone enanthate                    0.343800
         Transdermal gel          T                                         0.334286
         Transdermal gel + IM     Testosterone + TC                         0.333851
         Transdermal gel + SUBQ   Testosterone + TC                         0.324211
         Transdermal patch        T                                         0.335090
         Name: Peak hemoglobin/hematocrit on testosterone, dtype: float64
```

```
In [ ]:
```

% with each management type (none, lower dose or phlebotomy)

```
In [46]: df.groupby(['Route of administration', 'Testosterone type'])['Management (dose reduction,
          phlebotomy, etc)'].count()
```

```
Out[46]: Route of administration  Testosterone type
         IM                       Cypionate                             132
                                  Testosterone enanthate                  6
                                  Testosterone enanthate + Cypionate      2
         IM + SUBQ                Cypionate                               1
         IM + Transdermal patch   Cypionate                               2
                                  Cypionate + Testosterone                1
         Powder/cream             T                                       1
         SUBQ                     Cypionate                              29
                                  Testosterone enanthate                  4
         Transdermal gel          T                                      27
         Transdermal gel + IM     Testosterone + TC                       4
         Transdermal gel + SUBQ   Testosterone + TC                       1
         Transdermal patch        T                                       5
         Name: Management (dose reduction, phlebotomy, etc), dtype: int64
```

```
In [47]: df['Management (dose reduction, phlebotomy, etc)'].value_counts()
```

```
Out[47]: NONE            202
         Dose reduced     13
         Name: Management (dose reduction, phlebotomy, etc), dtype: int64
```

```
In [48]: df['Management (dose reduction, phlebotomy, etc)'].value_counts()/234 *100
```

```
Out[48]: NONE            86.324786
         Dose reduced     5.555556
         Name: Management (dose reduction, phlebotomy, etc), dtype: float64
```

```
In [ ]:
```

In [49]:

```
df.groupby(['Route of administration', 'Testosterone type','Management (dose reduction, phlebotomy, etc)'])['Management (dose reduction, phlebotomy, etc)'].count()/234*100
```

Out[49]:

```
Route of administration  Testosterone type                   Management (dose reduction,
phlebotomy, etc)
IM                       Cypionate                           Dose reduced
3.418803
                                                             NONE
52.991453
                         Testosterone enanthate              NONE
2.564103
                         Testosterone enanthate + Cypionate  NONE
0.854701
IM + SUBQ                Cypionate                           NONE
0.427350
IM + Transdermal patch   Cypionate                           NONE
0.854701
                         Cypionate + Testosterone            NONE
0.427350
Powder/cream             T                                   NONE
0.427350
SUBQ                     Cypionate                           Dose reduced
0.854701
                                                             NONE
11.538462
                         Testosterone enanthate              NONE
1.709402
Transdermal gel          T                                   Dose reduced
1.282051
                                                             NONE
10.256410
Transdermal gel + IM     Testosterone + TC                   NONE
1.709402
Transdermal gel + SUBQ   Testosterone + TC                   NONE
0.427350
Transdermal patch        T                                   NONE
2.136752
Name: Management (dose reduction, phlebotomy, etc), dtype: float64
```

In [50]:

```
df.to_csv(r'C:\freelancing project\newdf.xlsx')
```

Hgb Plot

In [51]:

```
import seaborn as sns
```

In [52]:

```
new = df[['Pre Hgb','Post Hgb','no_missing_column']].copy()
new
```

Out[52]:

	Pre Hgb	Post Hgb	no_missing_column
0	14.60	15.600000	non-missing
1	12.70	13.600000	non-missing
2	12.60	17.200001	non-missing
3	13.30	16.200001	non-missing
4	14.35	18.000000	non-missing
5	13.16	16.299999	non-missing
6	15.50	17.400000	non-missing
7	11.80	17.100000	non-missing
8	NaN	16.400000	non-missing
9	12.40	15.900000	non-missing
10	12.90	17.400000	non-missing
11	14.15	17.000000	non-missing
12	16.60	16.900000	non-missing
13	13.90	NaN	non-missing
14	NaN	NaN	non-missing
15	NaN	NaN	non-missing
16	13.30	NaN	non-missing
17	13.10	NaN	non-missing
18	13.20	NaN	non-missing
19	13.90	NaN	non-missing
20	13.10	NaN	non-missing
21	14.10	NaN	non-missing
22	13.23	NaN	non-missing
23	15.90	NaN	non-missing
24	14.30	NaN	non-missing
25	NaN	NaN	non-missing
26	13.00	NaN	non-missing
27	14.43	16.900000	non-missing
28	NaN	16.500000	non-missing
29	NaN	17.600000	non-missing
...	...	...	...
204	13.10	18.400000	non-missing
205	12.88	NaN	non-missing
206	11.80	NaN	non-missing
207	14.00	17.400000	non-missing
208	NaN	16.299999	non-missing
209	13.20	14.100000	non-missing
210	NaN	14.200000	non-missing
211	11.80	14.800000	non-missing
212	15.25	14.700000	non-missing

	Pre Hgb	Post Hgb	no_missing_column
213	12.95	15.000000	non-missing
214	NaN	13.200000	non-missing
215	12.30	15.400000	non-missing
216	NaN	17.100000	non-missing
217	14.70	17.200001	non-missing
218	13.83	16.900000	non-missing
219	14.45	16.700001	non-missing
220	14.64	13.700000	non-missing
221	15.80	NaN	non-missing
222	14.45	17.000000	non-missing
223	15.00	16.500000	non-missing
224	12.67	14.900000	non-missing
225	12.85	NaN	non-missing
226	12.30	13.500000	non-missing
227	NaN	12.800000	non-missing
228	NaN	19.700001	non-missing
229	NaN	NaN	non-missing
230	NaN	NaN	non-missing
231	NaN	NaN	non-missing
232	NaN	16.000000	non-missing
233	12.90	14.400000	non-missing

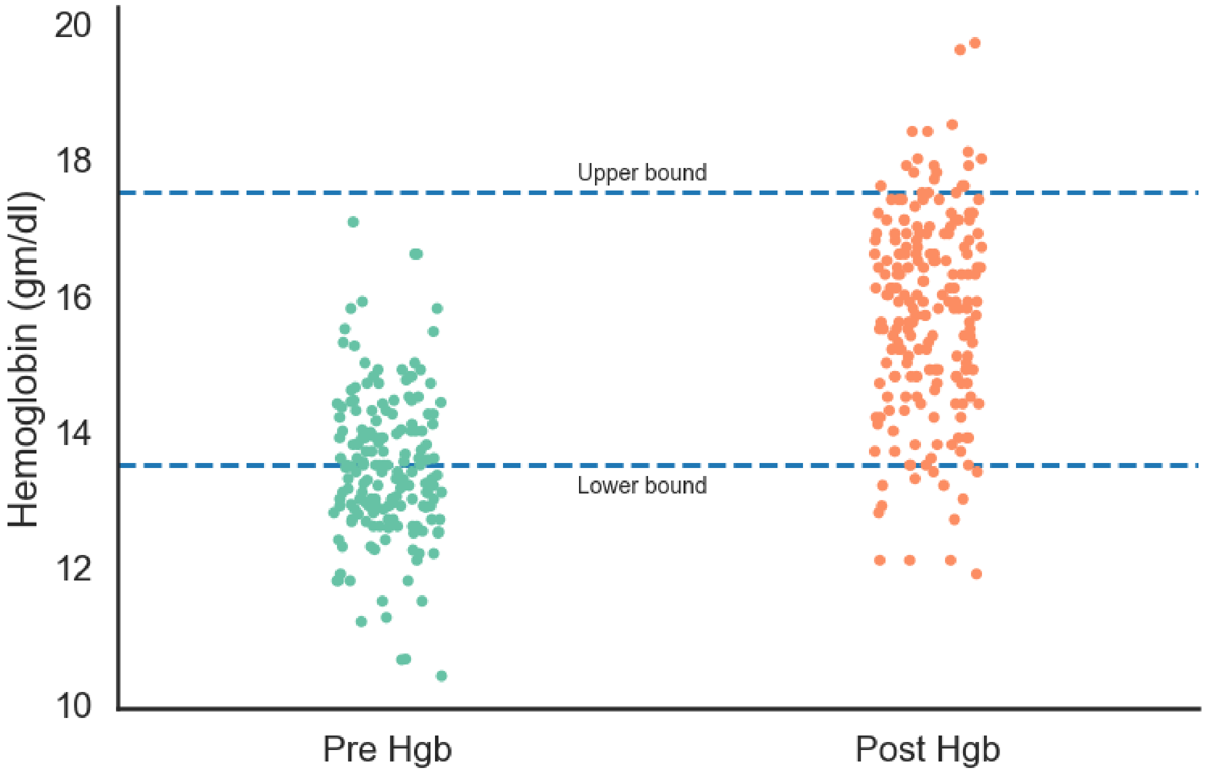
234 rows × 3 columns

In [53]:
```
new01 = df[['Pre Htc','Post Htc','no_missing_column']].copy()
```

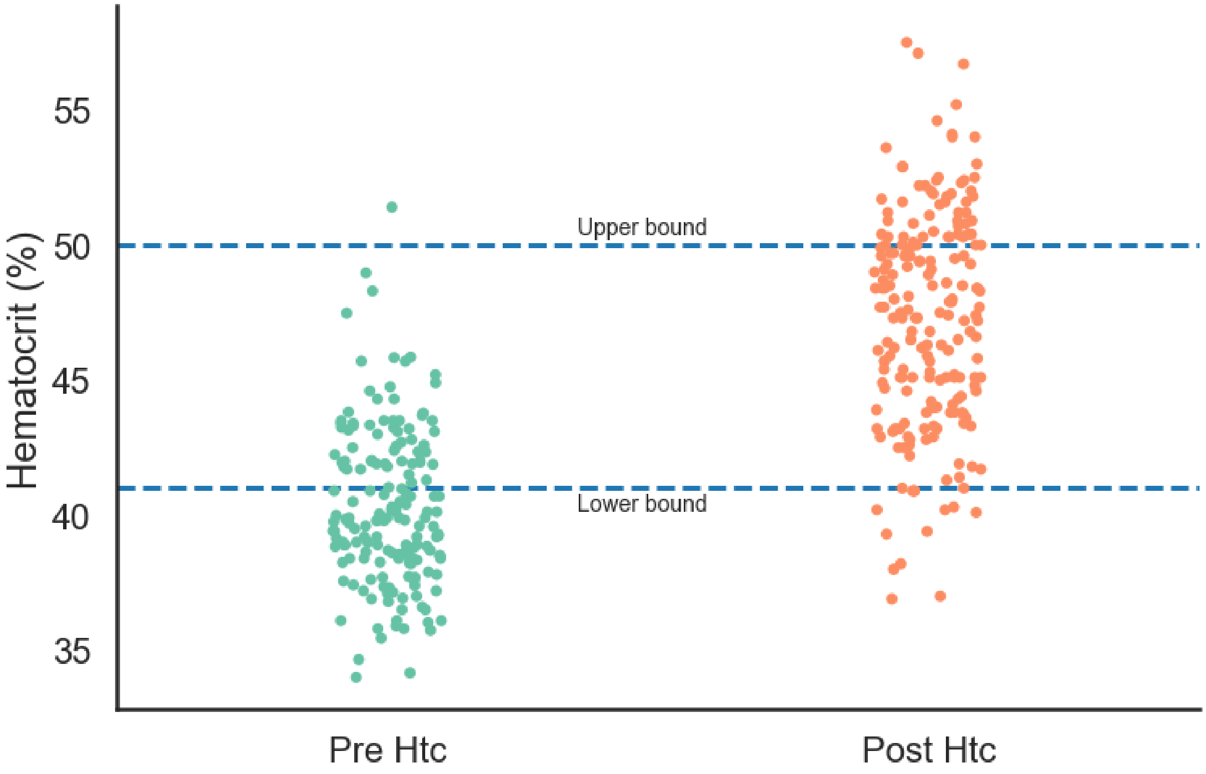
In [54]:
```
new=new.melt(id_vars='no_missing_column', var_name='Hgb', value_name='Value')
new01=new01.melt(id_vars='no_missing_column', var_name='Htc', value_name='Value')
```

In [62]:

```
import matplotlib.pyplot as plt
sns.set_style("white")
plt.figure(figsize=(12,8))
ax = sns.stripplot(x="Hgb", y="Value", data=new, jitter=1, palette="Set2", dodge=True, siz
e = 7)
ax.set(xlabel="", ylabel = "Hemoglobin (gm/dl)")
ax.axhline(13.5, ls='--')
ax.axhline(17.5, ls='--')
ax.text(.35,13.1, "Lower bound", fontsize=14)
ax.text(.35,17.7, "Upper bound", fontsize=14)

sns.set_context("poster")
sns.despine()
```



In [56]:
```
import matplotlib.pyplot as plt
sns.set_style("white")
plt.figure(figsize=(12,8))
ax1 = sns.stripplot(x="Htc", y="Value", data=new01, jitter=1, palette="Set2", dodge=True,
size = 7)
ax1.set(xlabel="", ylabel = "Hematocrit (%)")
ax1.axhline(41, ls='--')
ax1.axhline(50, ls='--')
ax1.text(.35,40.2, "Lower bound", fontsize=14)
ax1.text(.35,50.4, "Upper bound", fontsize=14)

sns.set_context("poster")
sns.despine()
```

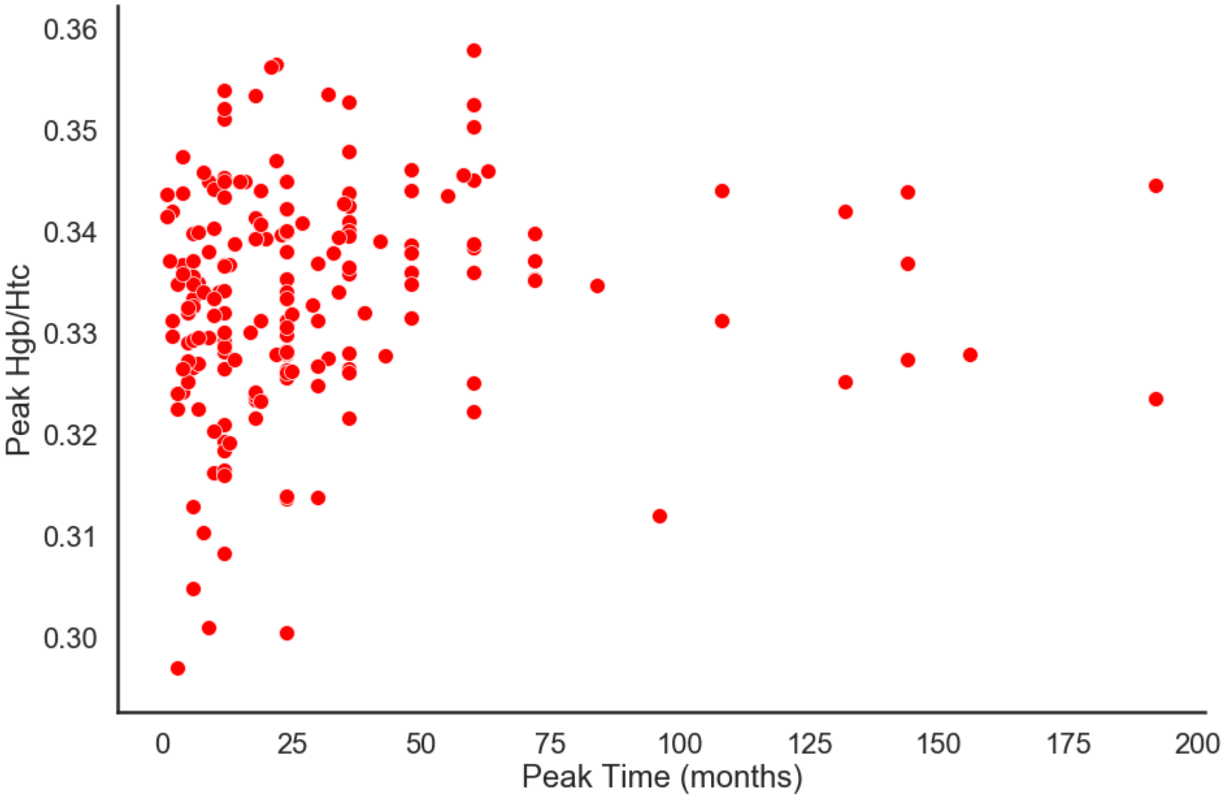


In [57]:
```
def group_age(age_df):
    if age_df <= 25:
        return "Age <= 25"
    elif age_df <= 35:
        return "25 < Age <= 35"
    else:
        return "Age > 35"

df["Age group"] = df['Age in years'].apply(group_age)
```

In [58]:
```
df["Age group"].value_counts()
```

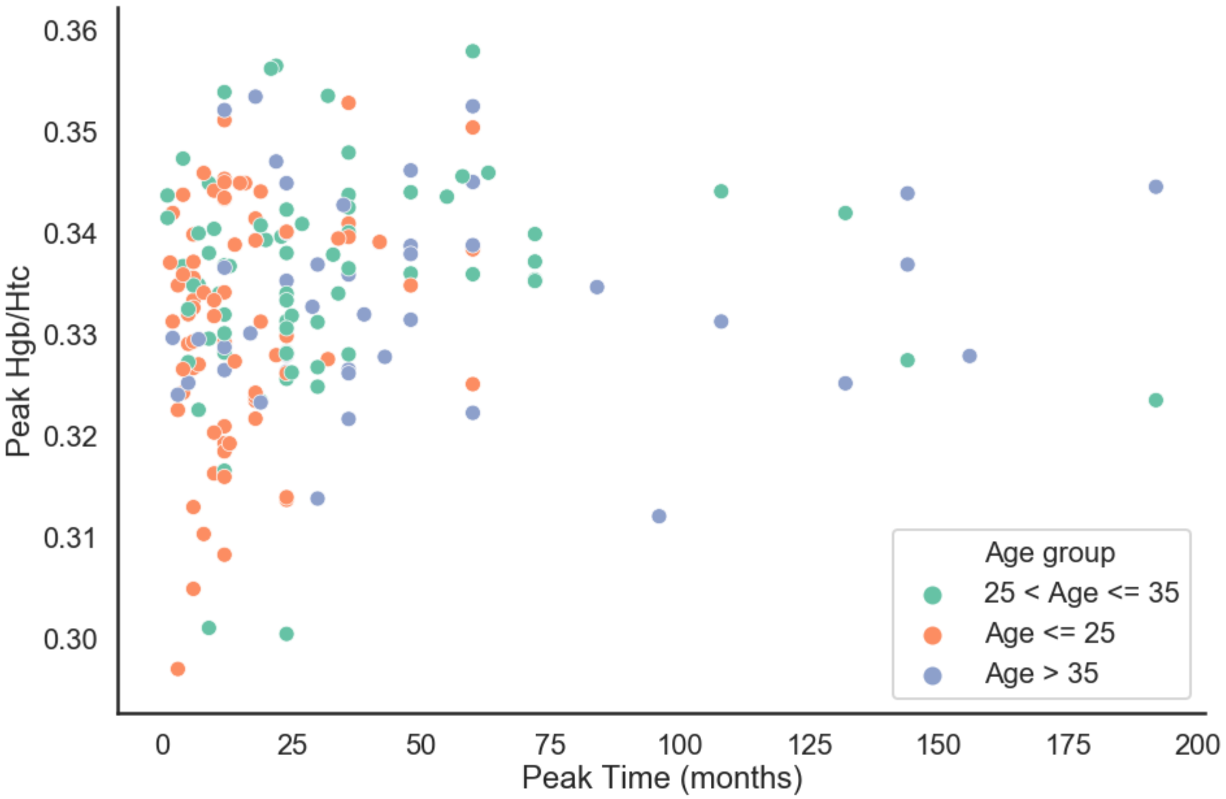
Out[58]:
```
25 < Age <= 35    94
Age <= 25         92
Age > 35          48
Name: Age group, dtype: int64
```

In [59]:
```
import matplotlib.pyplot as plt
sns.set_style("white")
plt.figure(figsize=(15,10))
ax02 = sns.scatterplot(x="When peak occurred", y="Peak hemoglobin/hematocrit on testostero
ne", data=df, palette ='Set2' ,color='red')
ax02.set(xlabel="Peak Time (months)", ylabel = "Peak Hgb/Htc")

sns.set_context("poster")
sns.despine()
```



In [60]:

```
import matplotlib.pyplot as plt
sns.set_style("white")
plt.figure(figsize=(15,10))
ax02 = sns.scatterplot(x="When peak occurred", y="Peak hemoglobin/hematocrit on testosterone",hue="Age group", data=df, palette ='Set2' ,color='red')
ax02.set(xlabel="Peak Time (months)", ylabel = "Peak Hgb/Htc")

sns.set_context("poster")
sns.despine()
```



In [ ]: