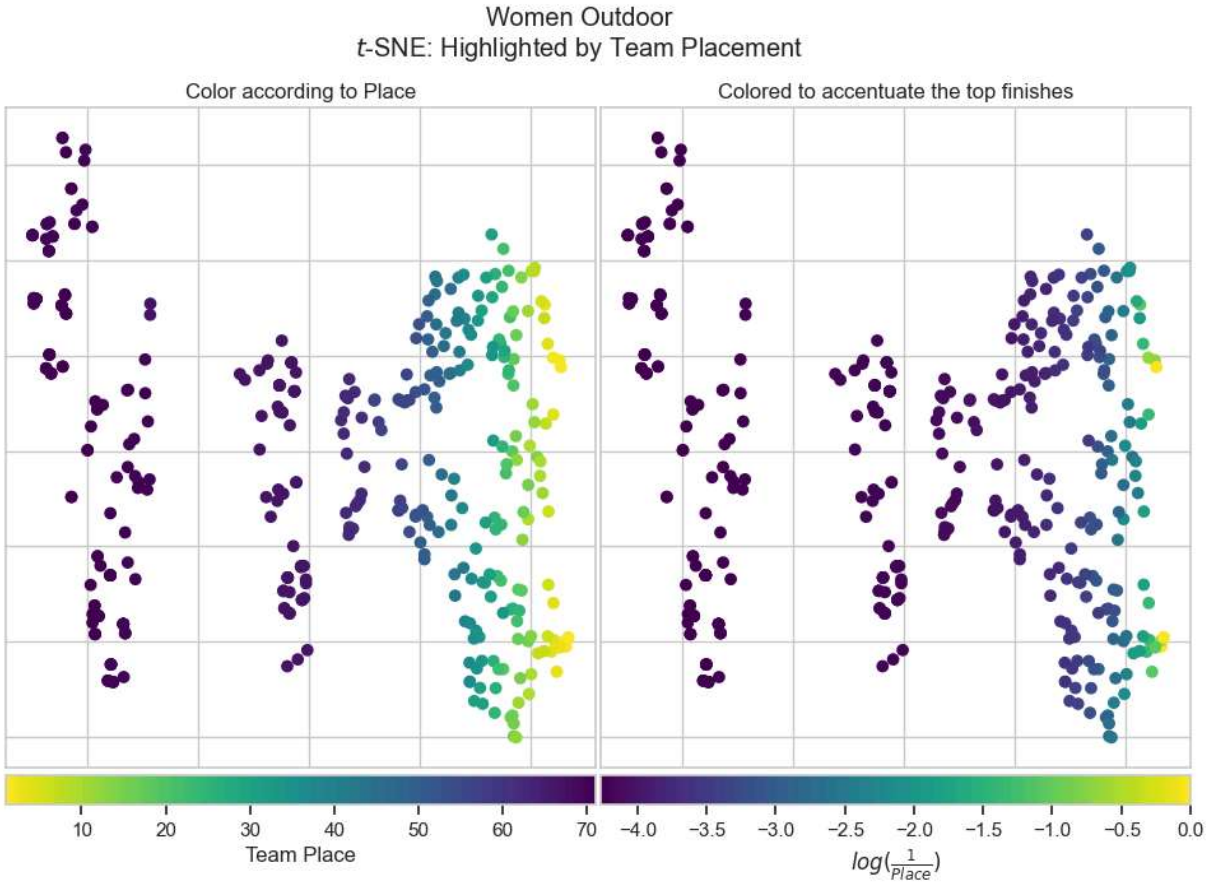
NCAA Track and Field Championships

Download Data and Add Metrics

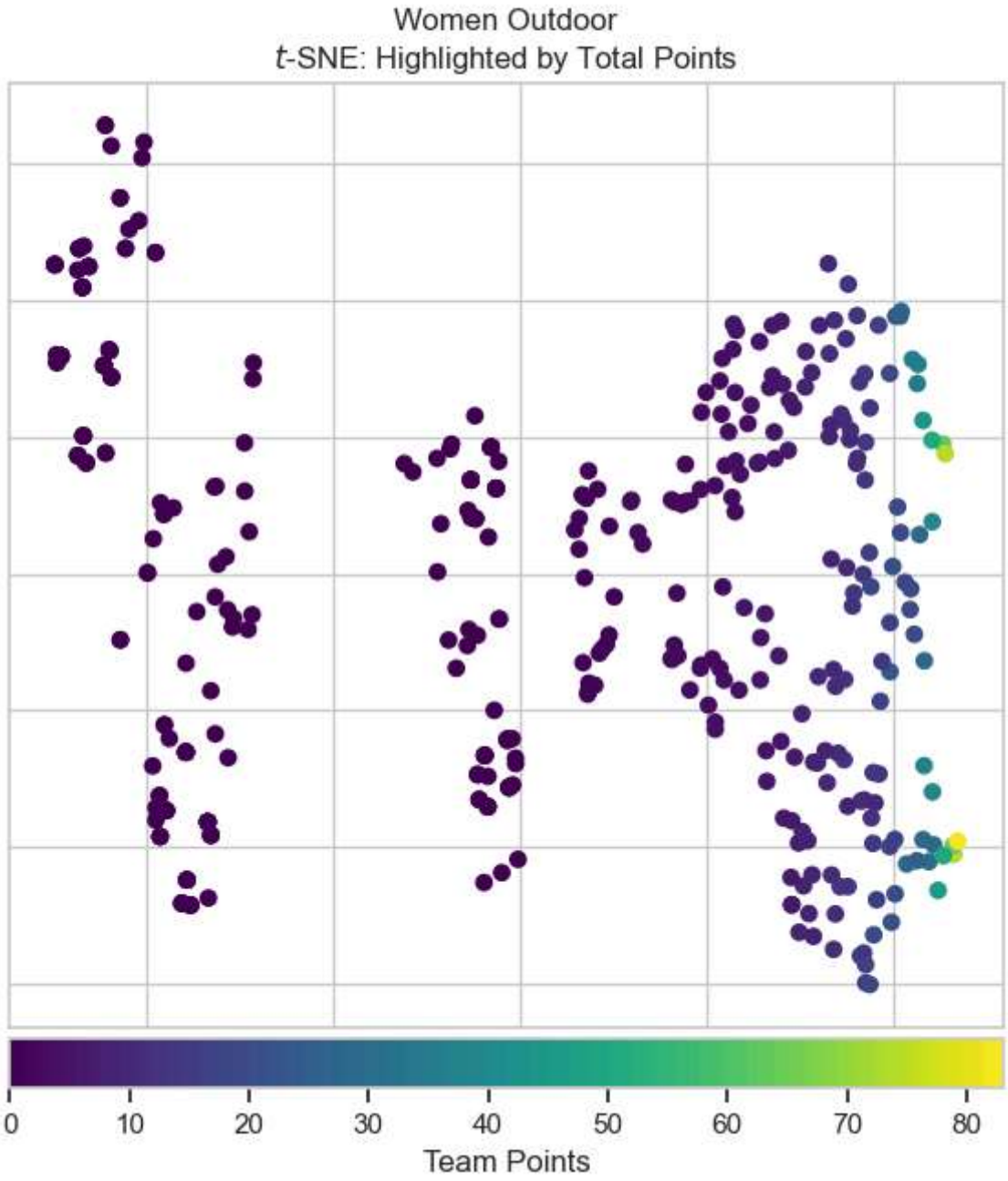
In [ ]:
```
dfs_f_o, dfs_m_o = download_flashresults(season='Outdoor', export=True)
dfs_f_i, dfs_m_i = download_flashresults(season='Indoor', export=True)

dfs_f_o = add_metrics(dfs_f_o)
```

Visualize Datasets

Women's Outdoor Championships

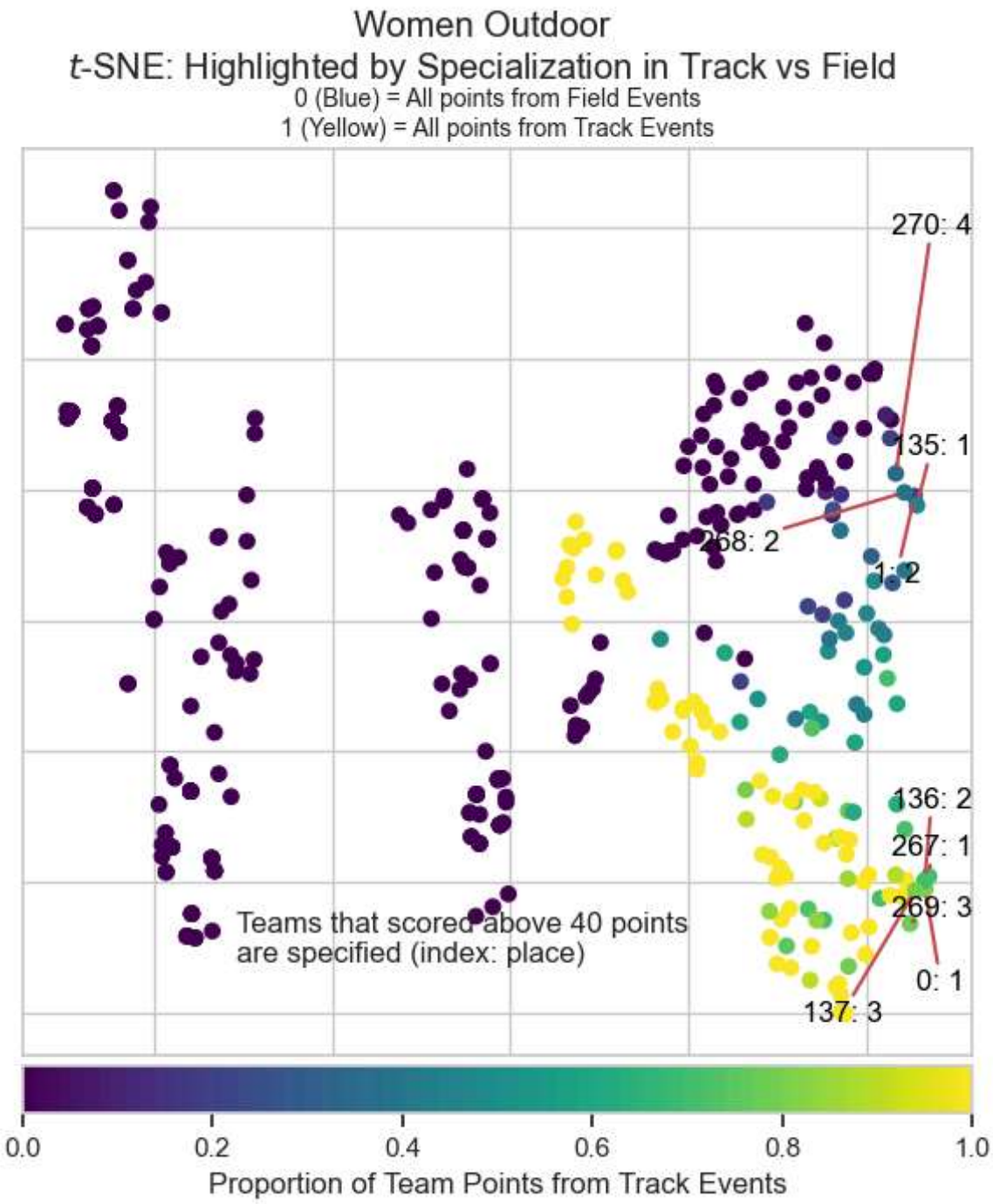


Scoring in the Top 10 seems to be possible in a variety of ways, but there are two distinct hubs for the top 3 places.



Again, we see there are clearly two distinct clusters that allow you to maximize points that will score you team podium finishes.

Specializing in Track or Field Points

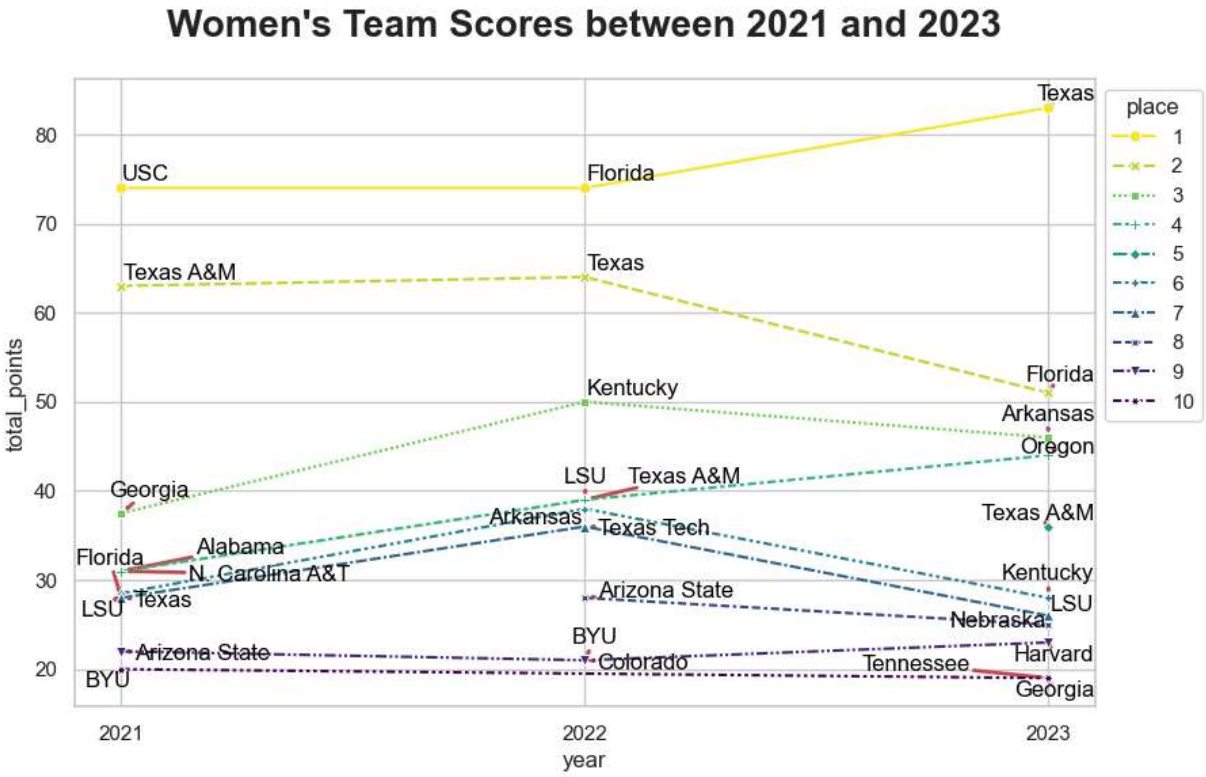


Clearly a strong degree of specialization, but both clusters of highest team points/placement involve points from both the track and the field.

Out[ ]:

	place	team_w	total_points	track_vs_field_track_proportion
267	1	Texas	83.00	0.71
0	1	USC	74.00	0.81
135	1	Florida	74.00	0.45
136	2	Texas	64.00	0.69
1	2	Texas A&M	63.00	0.22
268	2	Florida	51.00	0.39
137	3	Kentucky	50.00	0.80
269	3	Arkansas	46.00	0.78
2	3	Georgia	37.50	0.00

Team Points over Time

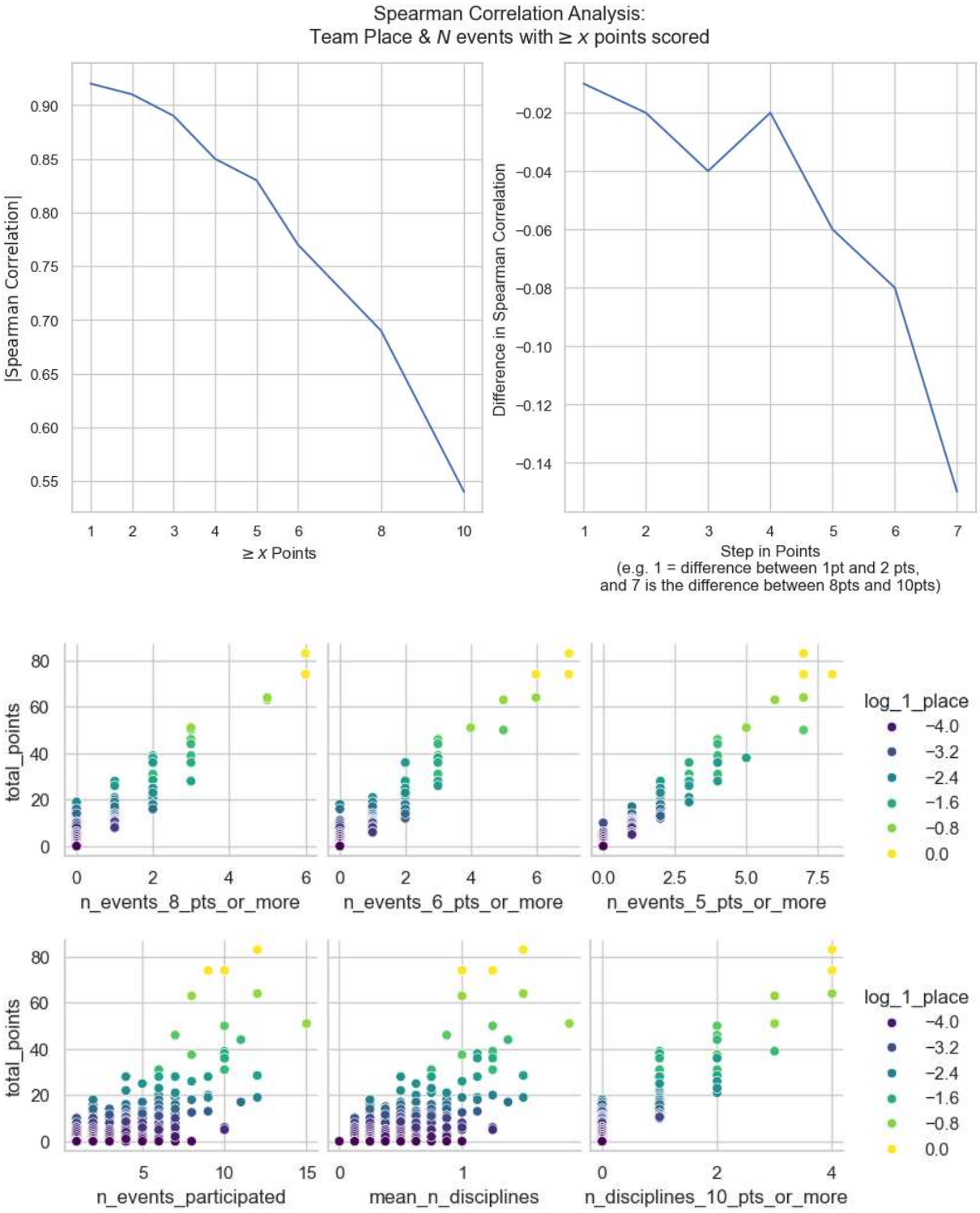


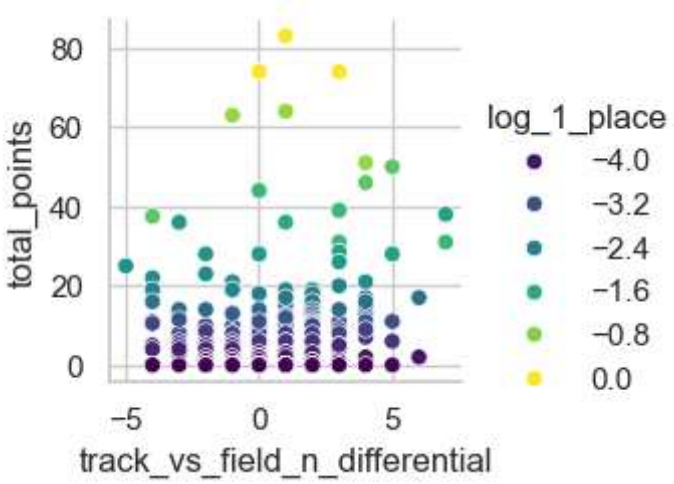
Correlations

Events

Spearman Correlation Coefficients

	total_points	place
n_events_participated	0.72	-0.68
n_events_10_pts_or_more	0.58	-0.54
n_events_8_pts_or_more	0.73	-0.69
n_events_6_pts_or_more	0.82	-0.77
n_events_5_pts_or_more	0.88	-0.83
n_events_4_pts_or_more	0.90	-0.85
n_events_3_pts_or_more	0.94	-0.89
n_events_2_pts_or_more	0.96	-0.91
n_events_1_pts_or_more	0.98	-0.92
n_events_sprints	0.43	-0.41
n_events_hurdles	0.36	-0.33
n_events_mid_distance	0.38	-0.37
n_events_distance	0.23	-0.22
n_events_relays	0.42	-0.41
n_events_jumps	0.32	-0.30
n_events_throws	0.34	-0.33
n_events_multi	0.27	-0.24
n_events_track	0.56	-0.53
n_events_field	0.50	-0.47
n_disciplines_10_pts_or_more	0.66	-0.62
mean_n_disciplines	0.72	-0.68


Spearman Correlation Analysis:
Team Place & N events with ≥ x points scored
|Spearman Correlation|
≥ x Points
Difference in Spearman Correlation
Step in Points
(e.g. 1 = difference between 1pt and 2 pts,
and 7 is the difference between 8pts and 10pts)
total_points
n_events_8_pts_or_more
n_events_6_pts_or_more
n_events_5_pts_or_more
log_1_place
n_events_participated
mean_n_disciplines
n_disciplines_10_pts_or_more



- Obviously, you want to participate in every event you can, and you want to score in every event you participate in. Additionally, you want to be scoring a minimum of 5-6 points in your events to have the best chances for winning. There are more track events than field events, so it makes sense that there's a higher correlation of track over field events. Relays have a relatively high impact, even though there's only three per championship event.
- There are more track events than field events
- Need to participate in around 7 events to be in the top 3.
- You want to score 10+ points in about 4 disciplines to have the best shot at winning.

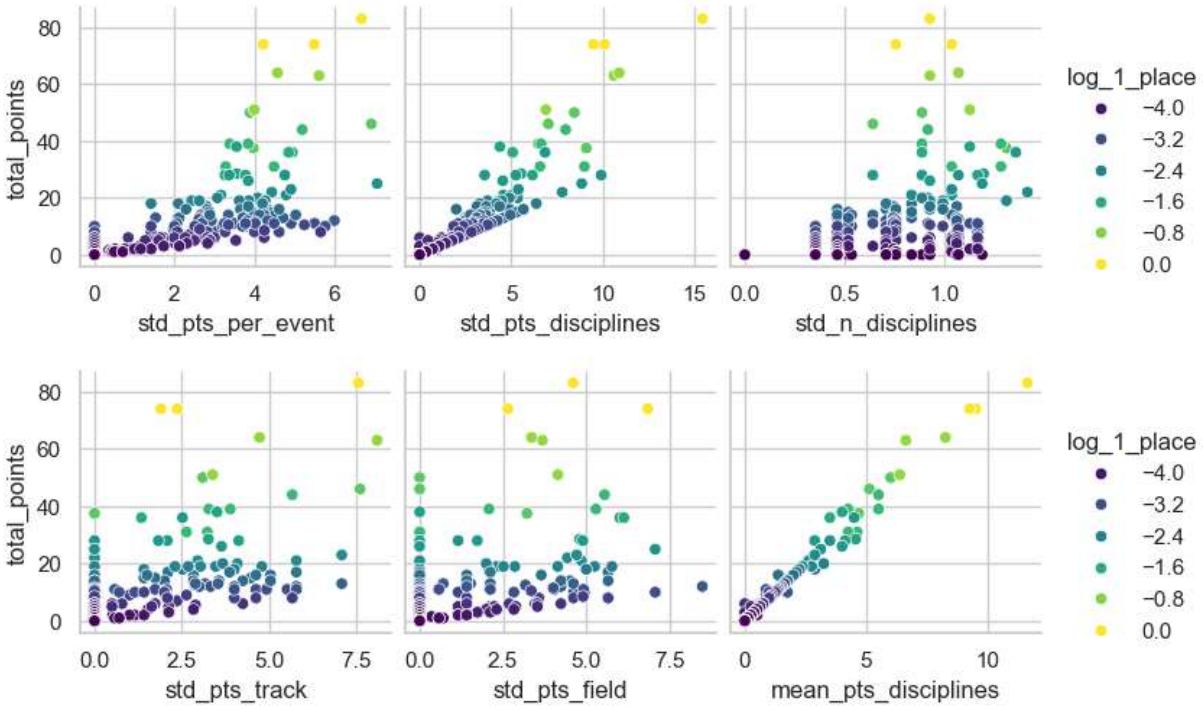
Points

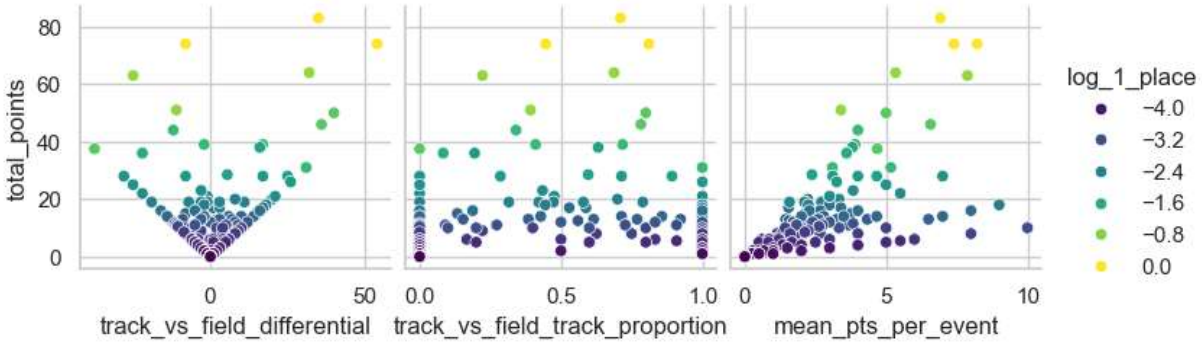
Higher standard deviations can indicate specialization/maxing out points in certain events while getting only a couple points in others. Where this fails as a metric is when you completely dominate a given discipline. For example, if you score 8+ points in each distance event, you'll have a smaller standard deviation.

Out[ ]:

Spearkman Correlation Coefficients

	total_points	place
std_pts_per_event	0.90	-0.85
std_pts_sprints	0.43	-0.41
std_pts_hurdles	0.24	-0.23
std_pts_mid_distance	0.31	-0.29
std_pts_distance	0.32	-0.29
std_pts_jumps	0.39	-0.37
std_pts_throws	0.33	-0.31
std_pts_track	0.70	-0.66
std_pts_field	0.62	-0.58
std_pts_disciplines	0.99	-0.94
std_n_disciplines	0.67	-0.63
mean_pts_disciplines	0.99	-0.94





- Specialization among disciplines is more important than specialization within a discipline.
- Need to score an average of 5 pts per discipline participated in to score in the top 3.
- If you're going to depend on field events, you really need to be dominant.