4.01-jk-eda-report

November 20, 2025

[1]:
```
# Ensure most recent (and not cached) source code is used
%load_ext autoreload
%autoreload 2
```

[2]:
```
# Import libraries
import json

import matplotlib.pyplot as plt
#plt.style.use("seaborn-v0_8-darkgrid")
plt.style.use("ggplot")

plt.rcParams['text.usetex'] = True
plt.rcParams['font.family'] = 'serif'
%matplotlib inline

import os

import pandas as pd
pd.options.display.max_columns = None
import scipy
import seaborn as sns
```

[3]:
```
# Import functions from other python files in the project
from pdga_score_scraper import *
from pdga_score_scraper.config import PROCESSED_DATA_DIR
```

```
2025-11-20 10:20:18.066 | INFO     |
pdga_score_scraper.config:<module>:11 - PROJ_ROOT

path is: /home/jon/repos/pdga-score-scraper
```

[4]:
```
# Set event_id from environment variable

event_id = os.getenv("NB_ARGS")
if event_id == None:
    event_id = 95560
    print(f"Environment variable not found.  Hardcoded as {event_id}.")
```

```
else:
    print(f"Processing data for event {event_id}.")
```

```
Processing data for event 95560.
```

[5]:
```
df = pd.read_json(PROCESSED_DATA_DIR / f"{event_id}-processed.json")

with open(PROCESSED_DATA_DIR / f"{event_id}-metadata-processed.json", 'r') as f:
    metadata = json.load(f)
```

[6]:
```
df.head()
```

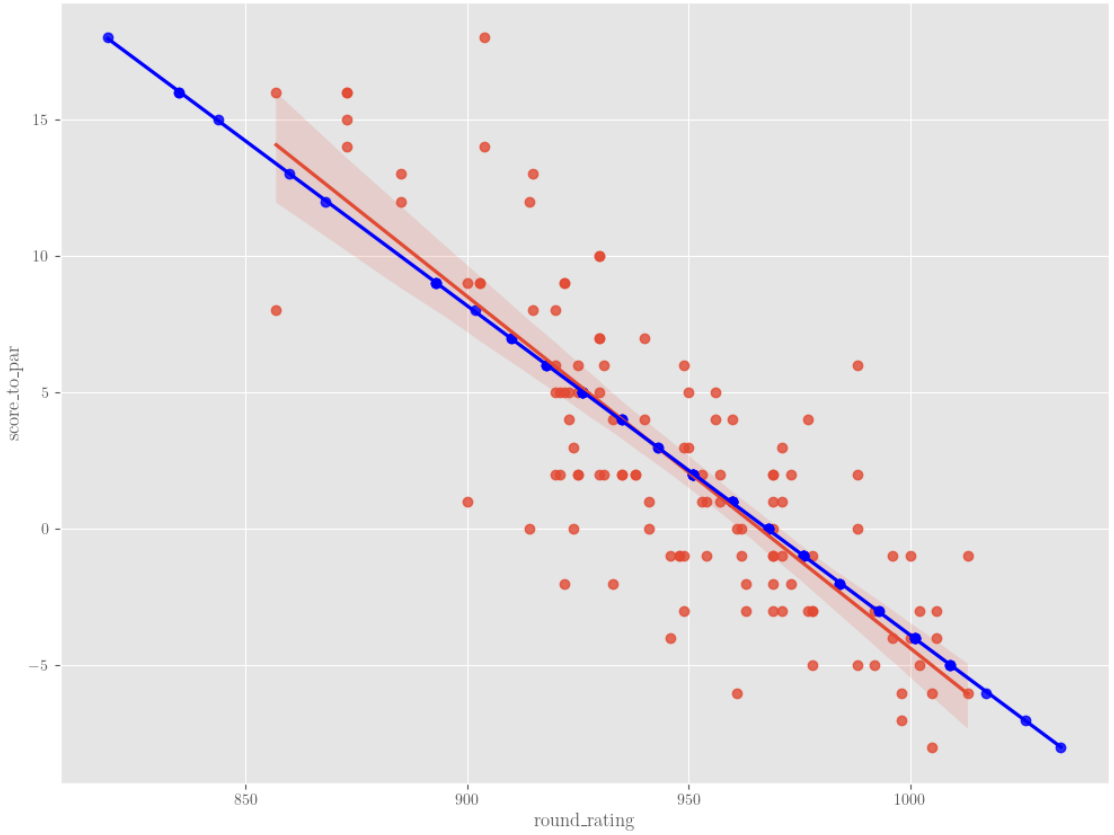
[6]:

	first_name	last_name	city	state	country	nationality	pdga_num	player_rating	rating_effective_date	player_class	result_id	tournament_division	tournament_place	tied_finish	formatted_place	shot_total	tournament_total_to_par	prize_usd	tournament_dnf	tournament_average_rating	tournament_rating_delta	score_id	round
0	Steve	Brinster	Warwick	NY	US	None	10628	1005	1760400000000	Professional	212107694	MPO	1	False	1st	110	-14	588	False	1026	21	25985480	1
1	Steve	Brinster	Warwick	NY	US	None	10628	1005	1760400000000	Professional	212107694	MPO	1	False	1st	110	-14	588	False	1026	21	26000869	2
2	Mike	Carman	Ottsville	PA	US	None	48559	998	1762819200000	Professional	212040498	MPO	2	False	2nd	111	-13	406	False	1022	24	25986622	1
3	Mike	Carman	Ottsville	PA	US	None	48559	998	1762819200000	Professional	212040498	MPO	2	False	2nd	111	-13	406	False	1022	24	26000867	2
4	Dylan	Horst	West Chester	PA	US	None	32599	1013	1760400000000	Professional	212089092	MPO	6	True	T-6th	117	-7	171	False	997	-16	25985549	1

```
             tee_time    course_name                           layout_name  \
0 2025-11-20 09:15:00  Camp Tecumseh  Tecumseh Takeover 2025 Pro Day Layout
1 2025-11-20 13:15:00  Camp Tecumseh  Tecumseh Takeover 2025 Pro Day Layout
2 2025-11-20 09:15:00  Camp Tecumseh  Tecumseh Takeover 2025 Pro Day Layout
3 2025-11-20 13:15:00  Camp Tecumseh  Tecumseh Takeover 2025 Pro Day Layout
4 2025-11-20 09:15:00  Camp Tecumseh  Tecumseh Takeover 2025 Pro Day Layout

   par  round_score  score_to_par  round_dnf  num_birdies  num_bogeys  \
0   62           54            -8      False           11           3
1   62           56            -6      False            7           1
2   62           55            -7      False            8           1
3   62           56            -6      False            8           2
4   62           56            -6      False            7           1

   round_rating  num_holes  layout_length  Hole 1  Hole 2  Hole 3  Hole 4  \
0          1034         18           8766       4       3       5       3
1          1017         18           8766       3       4       3       5
2          1026         18           8766       3       4       2       4
3          1017         18           8766       3       3       3       4
4          1017         18           8766       3       4       3       3

   Hole 5  Hole 6  Hole 7  Hole 8  Hole 9  Hole 10  Hole 11  Hole 12  Hole 13  \
0       2       4       2       4       2        3        3        2        3
1       2       4       2       4       3        3        3        3        3
2       3       5       2       3       3        3        3        2        3
3       2       4       3       4       2        3        3        2        3
4       2       3       3       4       2        4        3        3        3

   Hole 14  Hole 15  Hole 16  Hole 17  Hole 18
0        3        3        3        2        3
1        2        3        3        2        4
2        2        3        5        2        3
3        3        3        3        2        6
4        3        2        3        4        4
```

[7]:
```
fig, axs = plt.subplots(1, 1, figsize = (12, 9))

sns.regplot(data=df, x="player_rating", y = "score_to_par")
#ax.set_title("Monkies")
#ax.set_xlabel("PDGA player rating")
#ax.set_ylabel("Round score relative course par")

sns.regplot(data=df[df['round'] == 1], x="round_rating", y = "score_to_par",␣
 ↪color="blue")

#ax.legend()
```

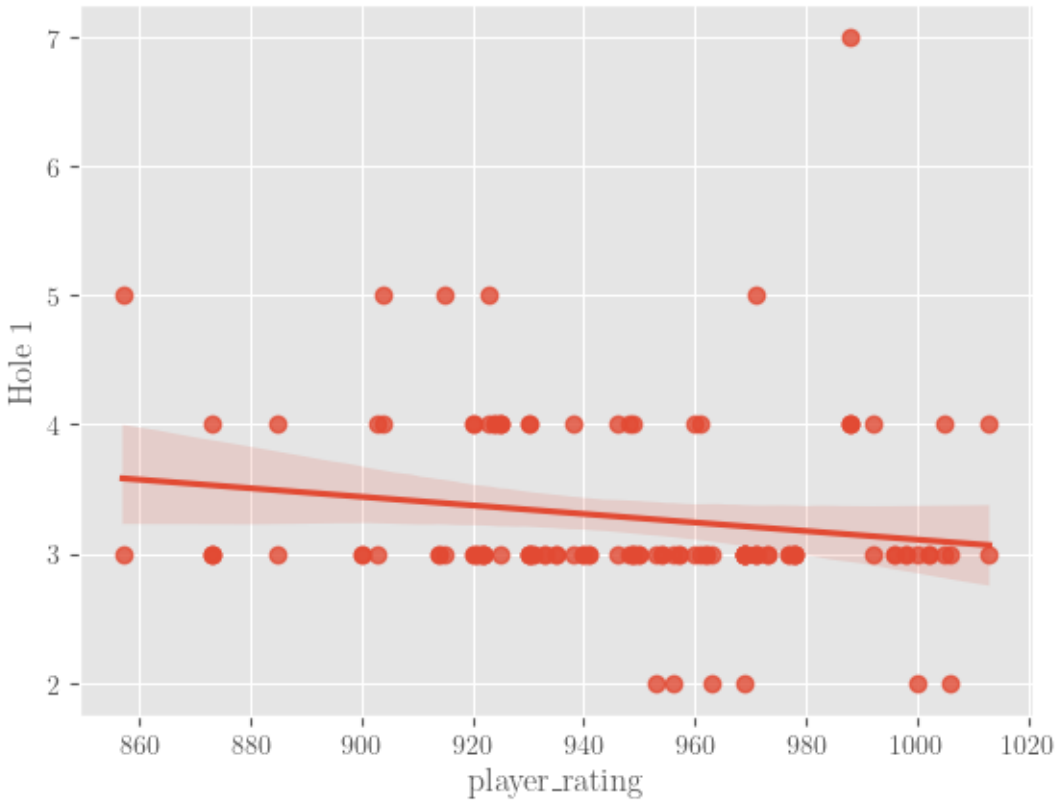
```
plt.show()
```

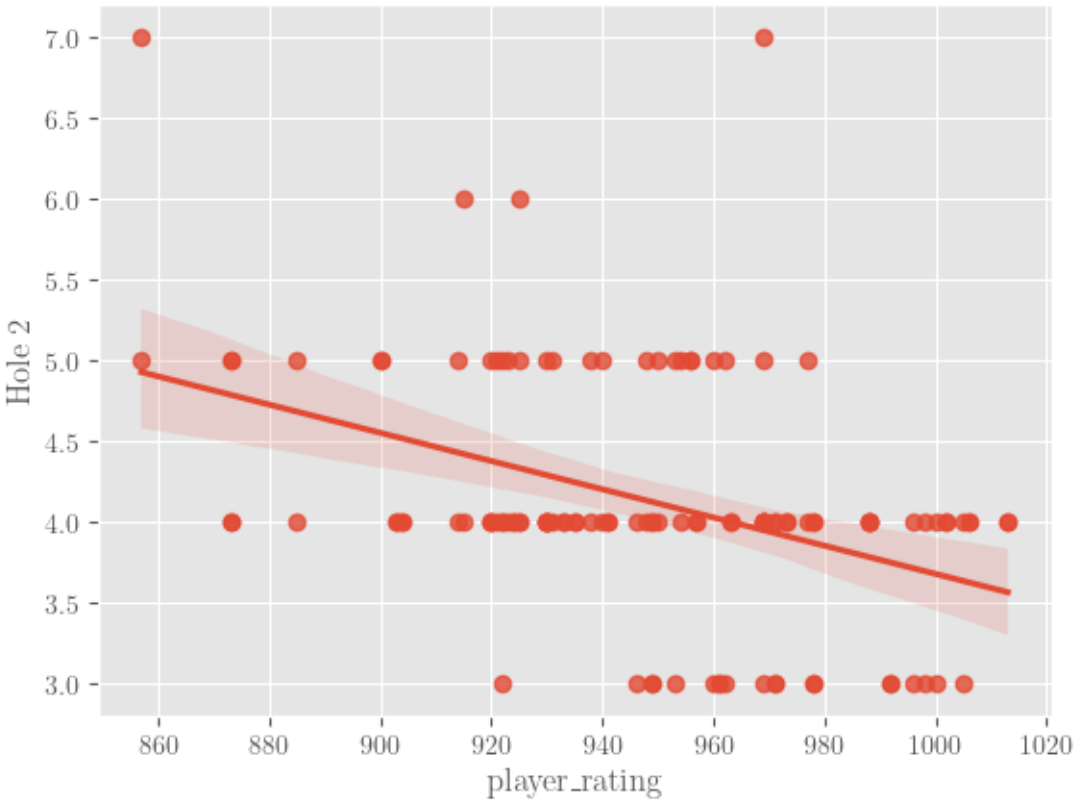


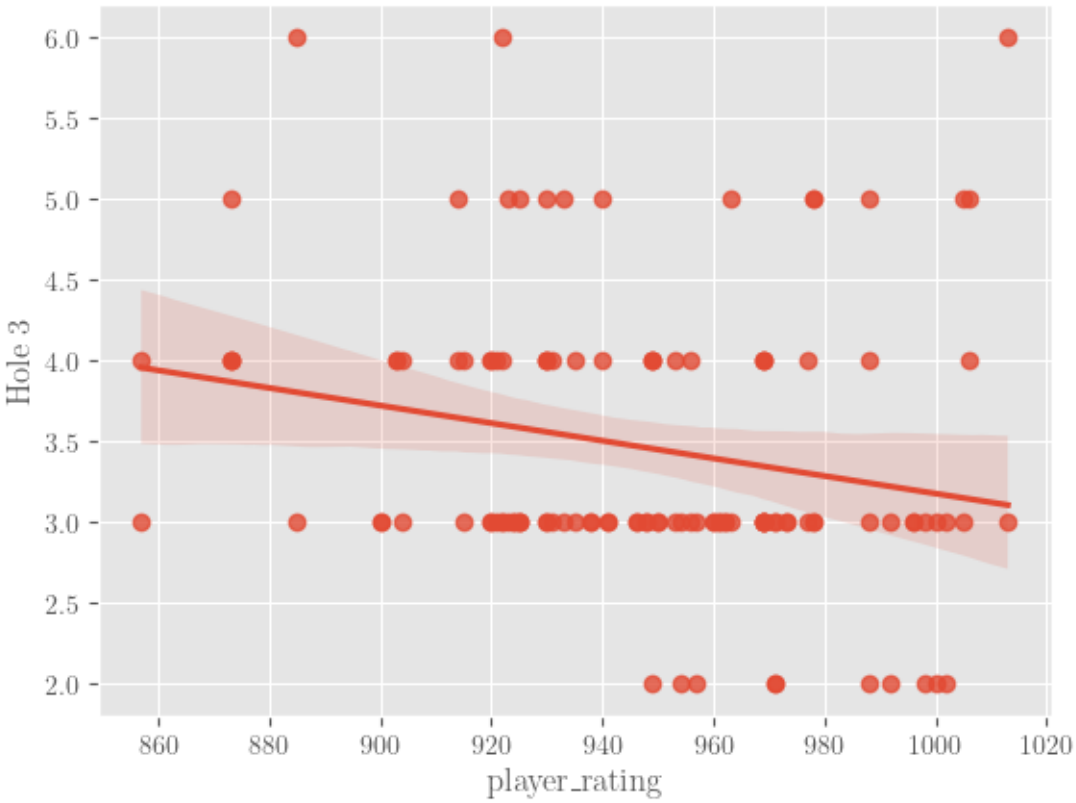
0.1 Regressing individual hole scores against player rating

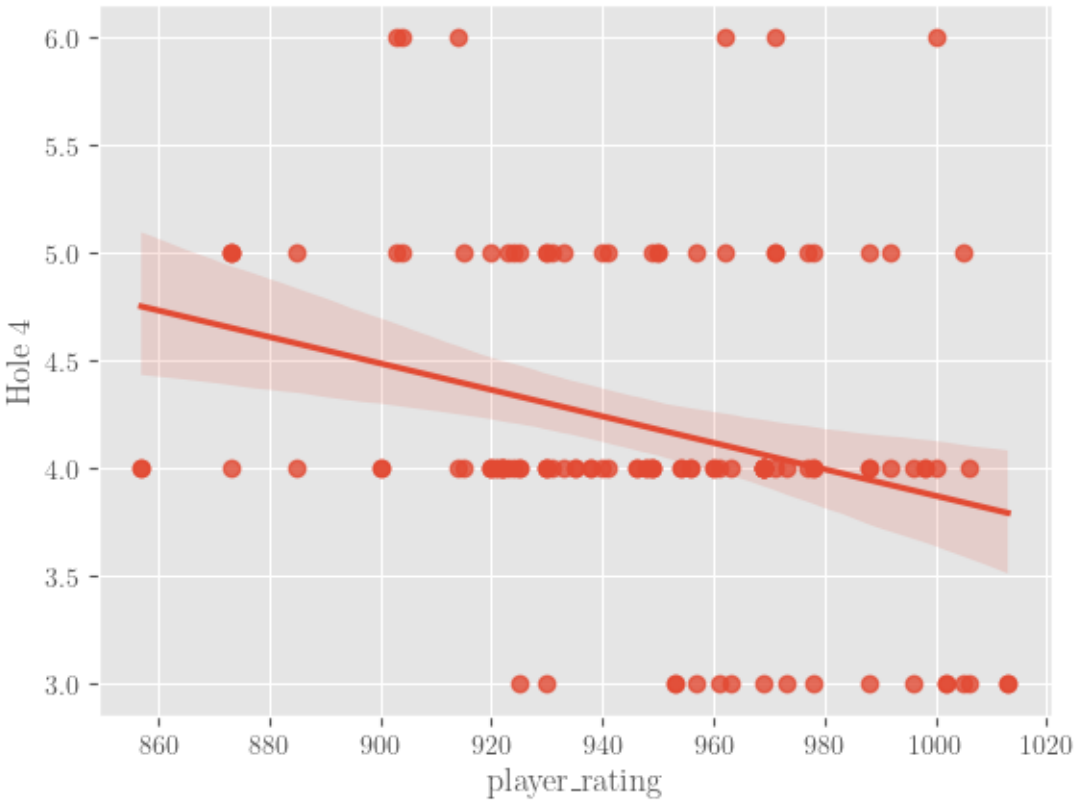
Does hole par inherently affect regresssion slope? Guessing yes. Maybe compare only equally-parred holes to evaluate differentiation.

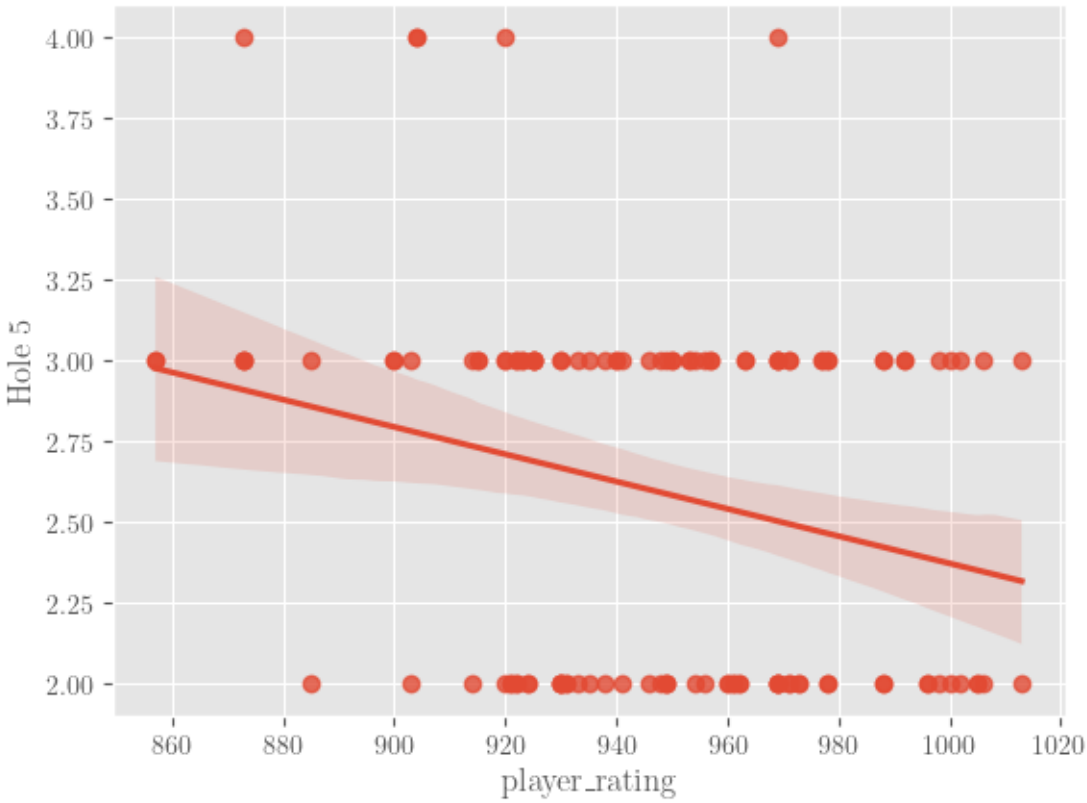
[8]:
```
slopes = {}
for i in range(1, 19):
    p = sns.regplot(data=df, x="player_rating", y = f"Hole {i}")

    slope, intercept, r, p, sterr = scipy.stats.linregress(x=p.get_lines()[0].
 ↪get_xdata(),
                                                           y=p.get_lines()[0].
 ↪get_ydata())
    slopes[i] = slope
    plt.show()
```

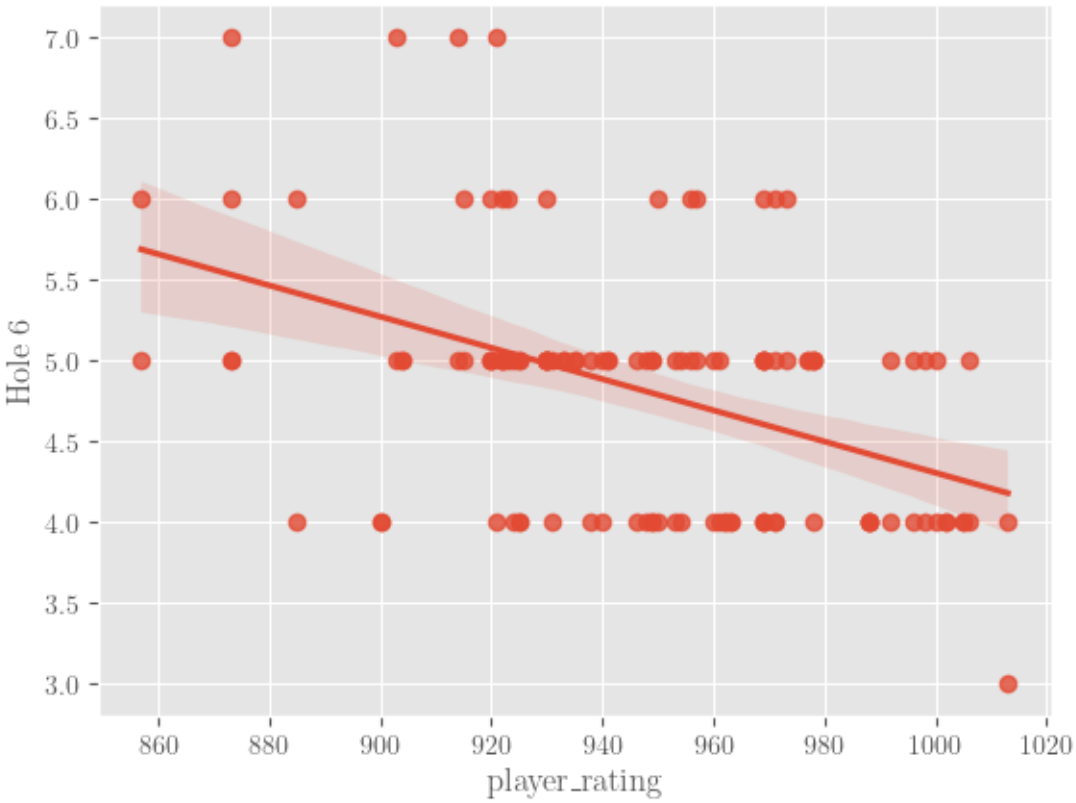

Hole 1
7
6
5
4
3
2
860
880
900
920
940
960
980
1000
1020
player_rating

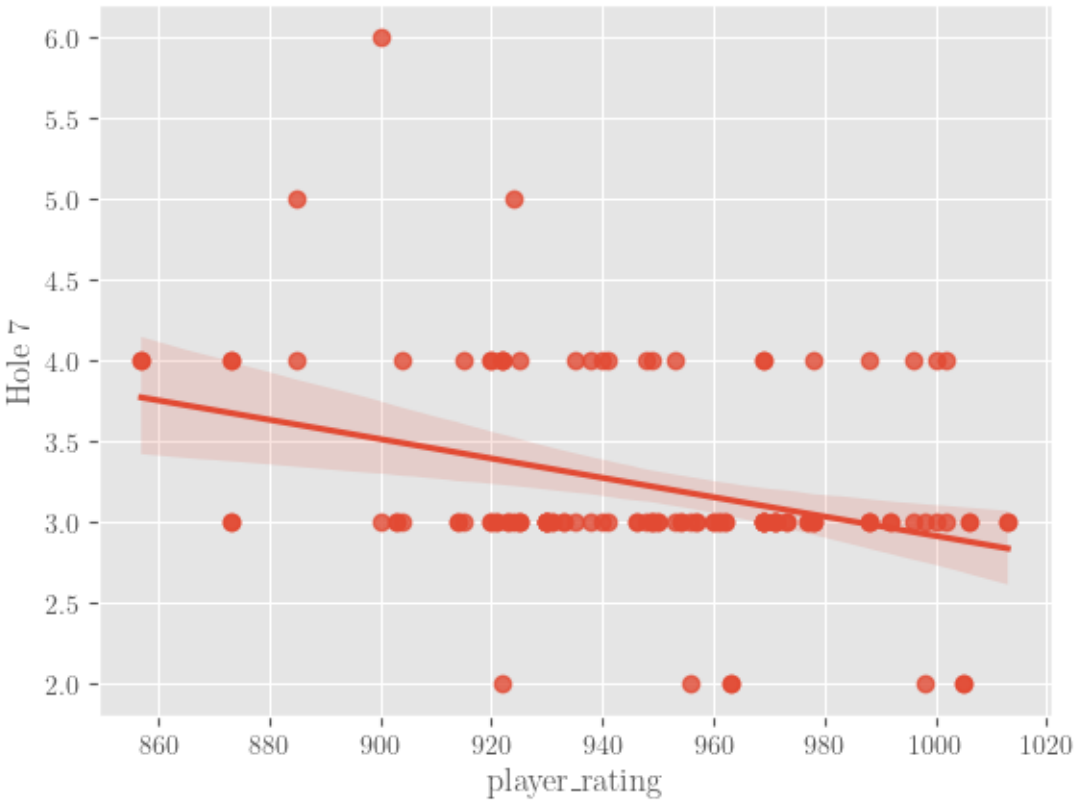

Hole 2
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
860
880
900
920
940
960
980
1000
1020
player_rating

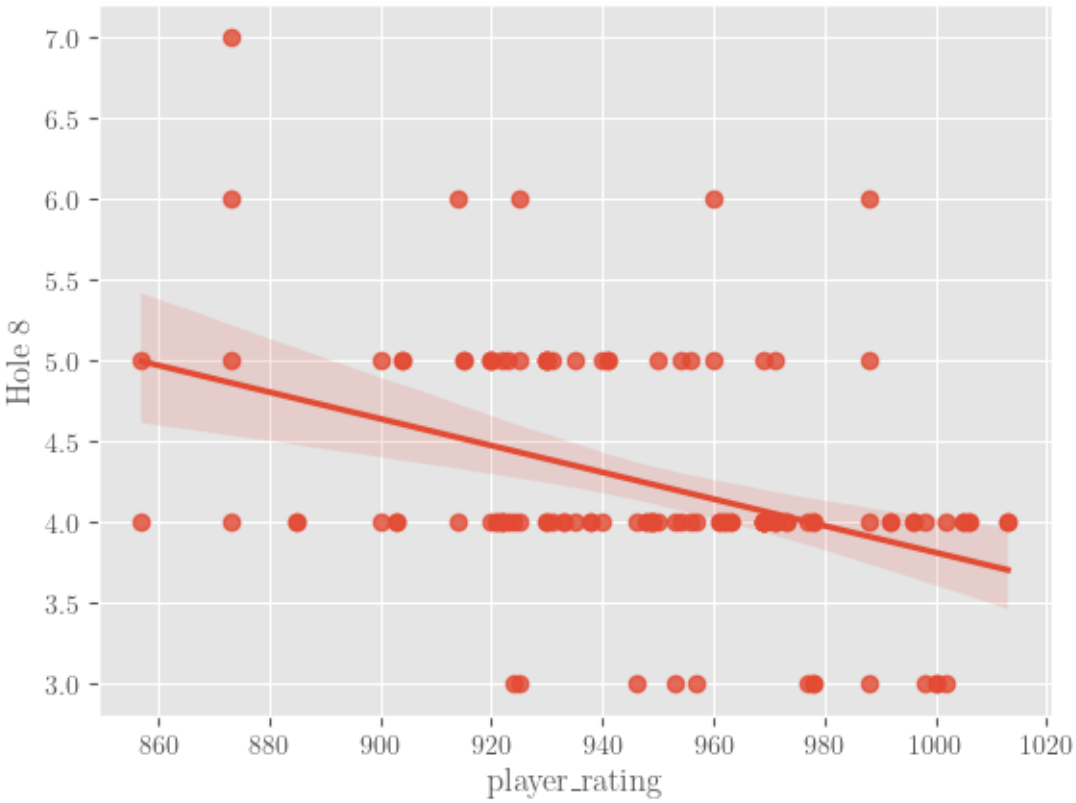

6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
Hole 3
860
880
900
920
940
960
980
1000
1020
player_rating

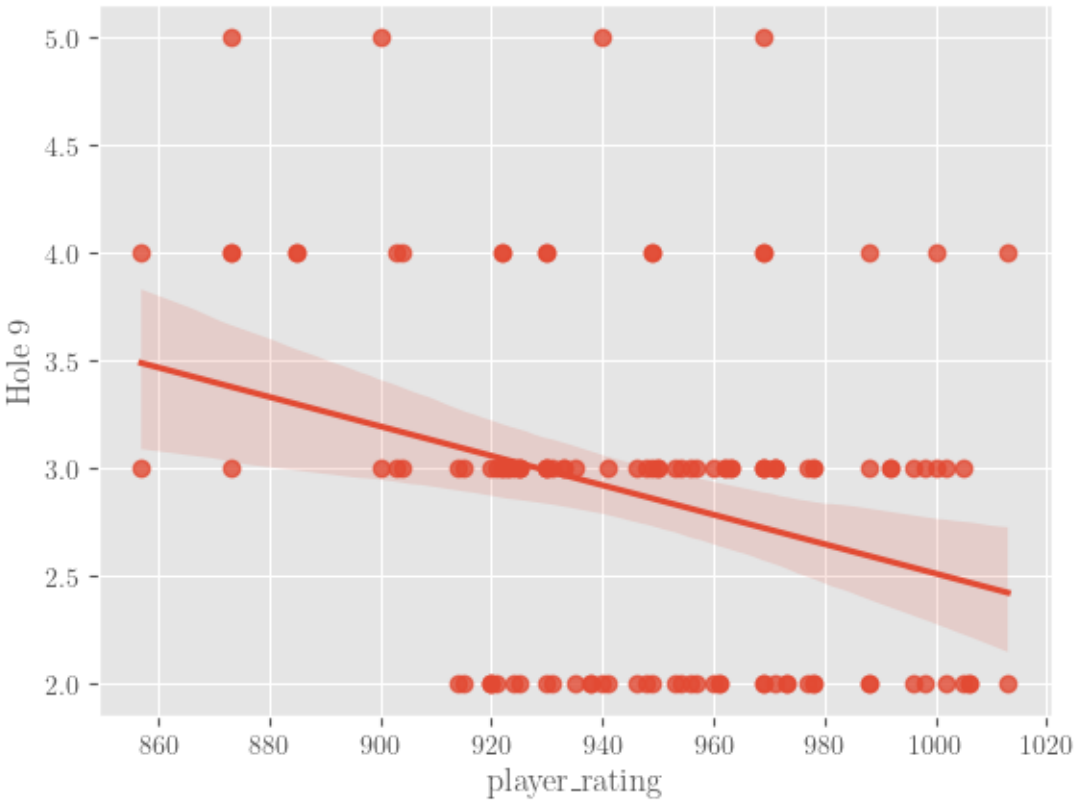

6.0
5.5
5.0
4.5
4.0
3.5
3.0
Hole 4
860
880
900
920
940
960
980
1000
1020
player_rating

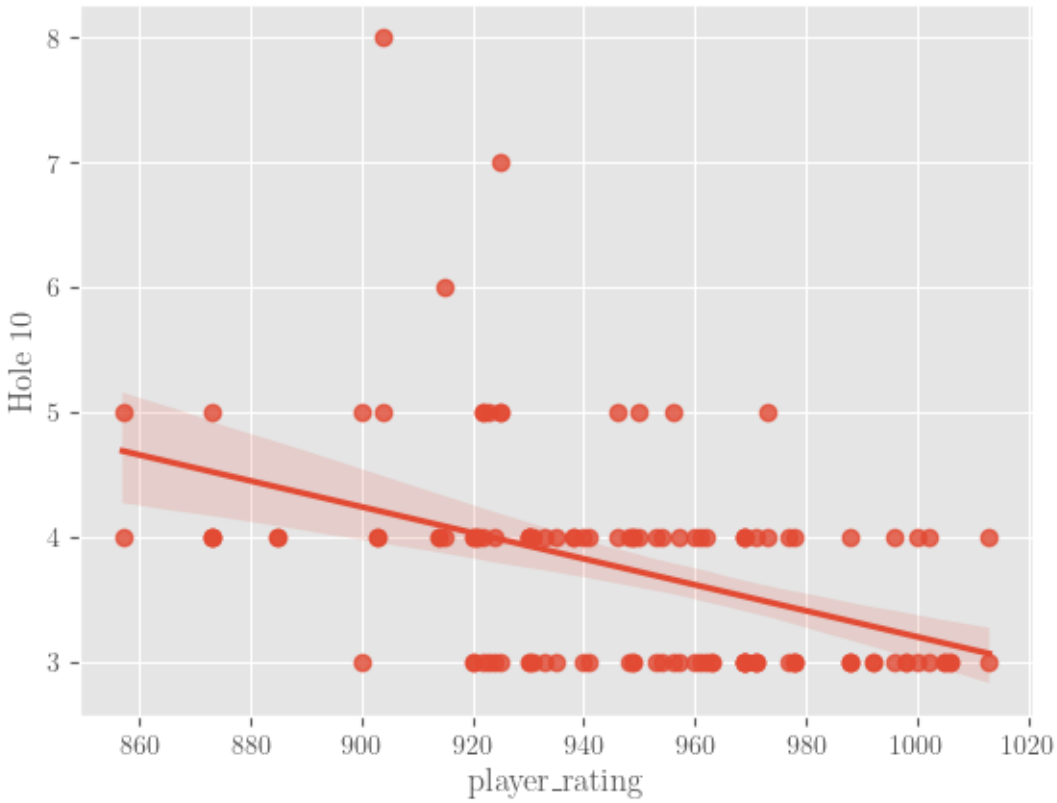

4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
Hole 5
860
880
900
920
940
960
980
1000
1020
player_rating

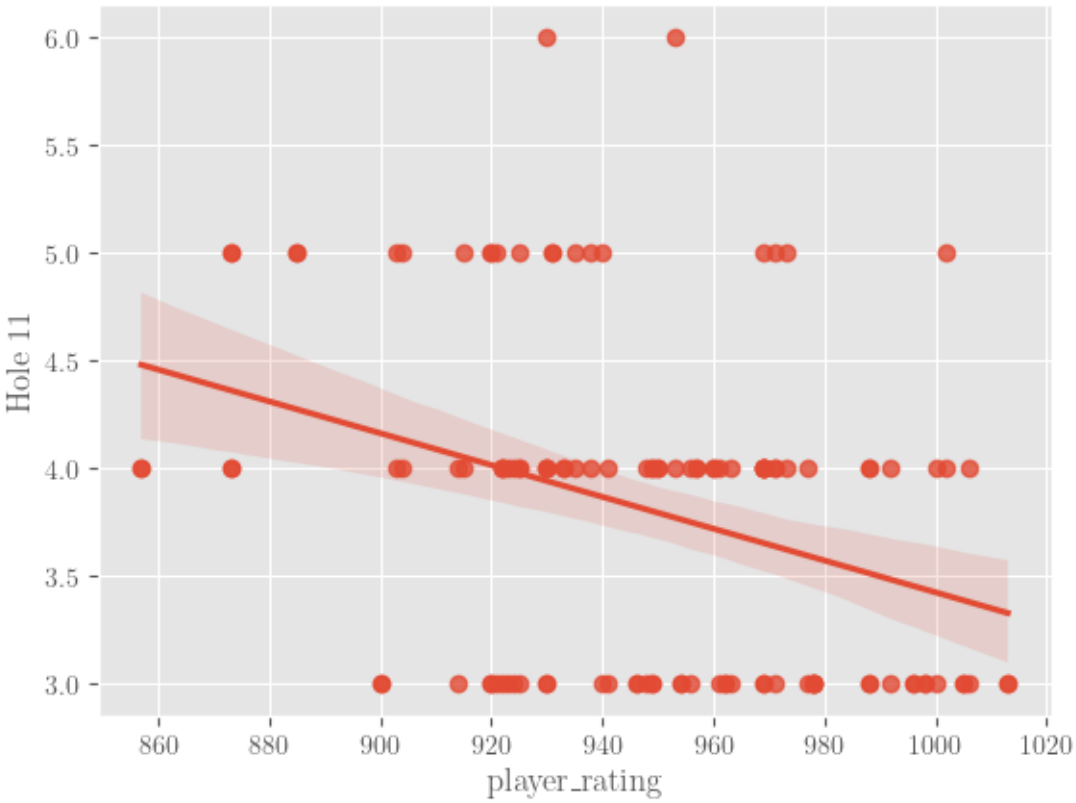

Hole 6
player_rating
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
860
880
900
920
940
960
980
1000
1020

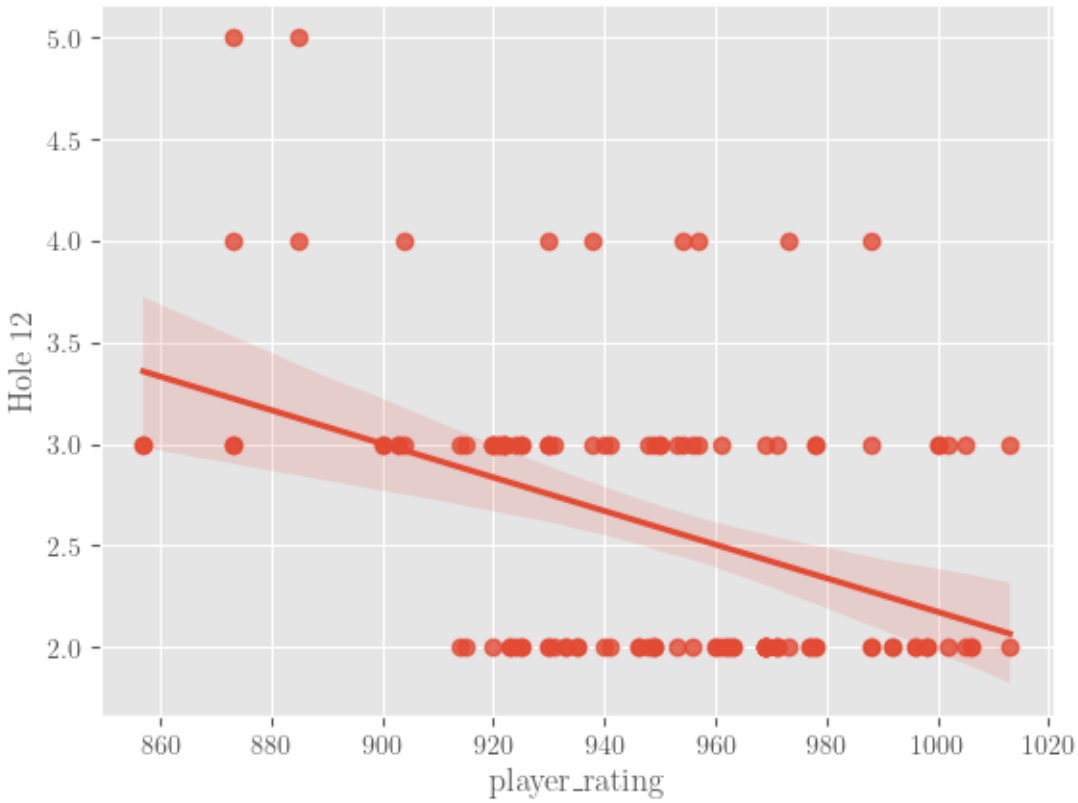

6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
Hole 7
860
880
900
920
940
960
980
1000
1020
player_rating

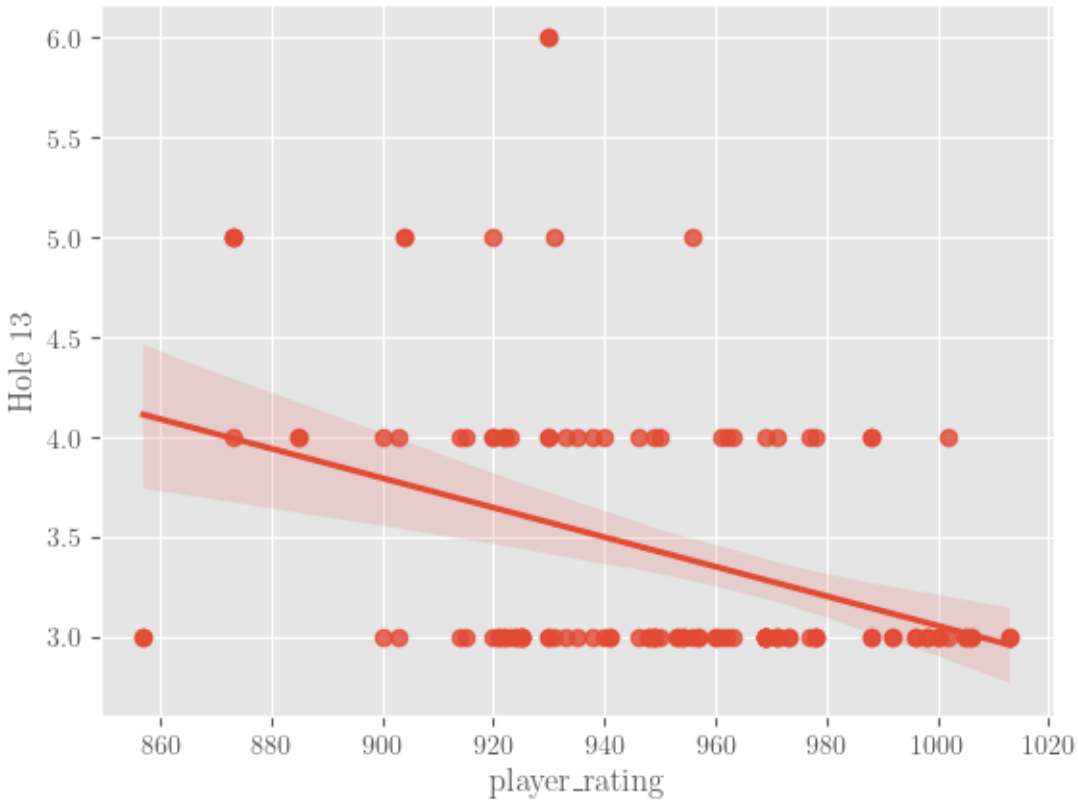

7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
Hole 8
860
880
900
920
940
960
980
1000
1020
player_rating

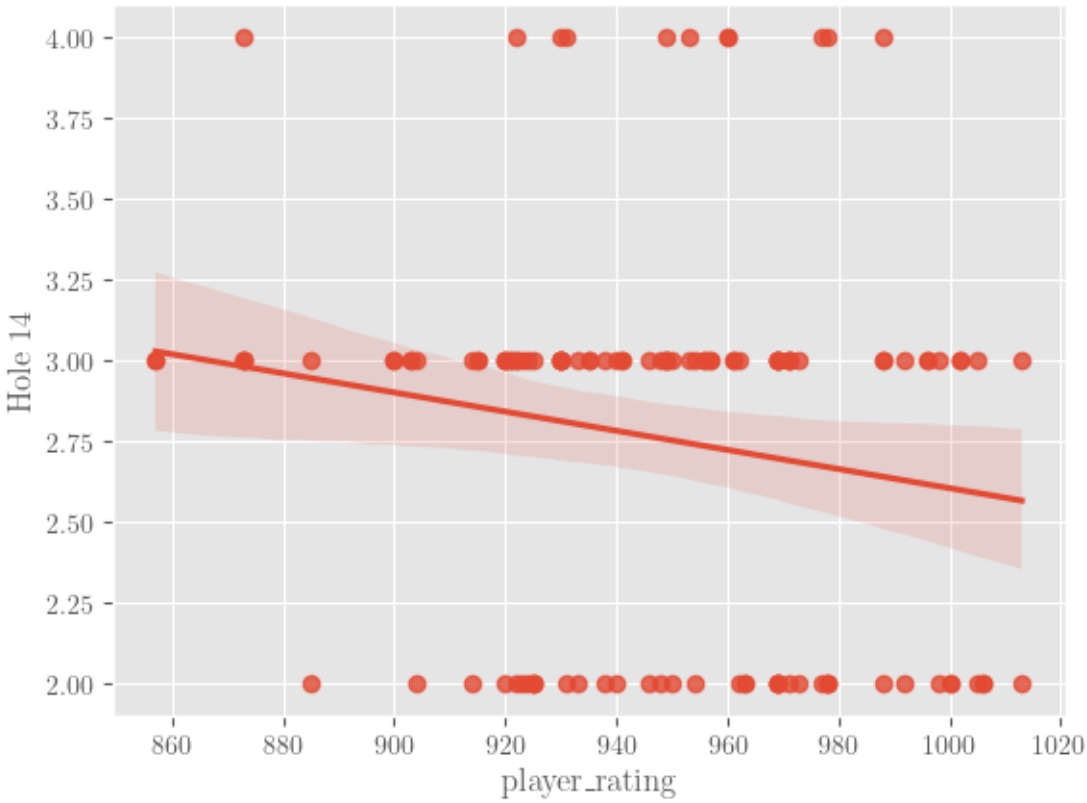

5.0
4.5
4.0
3.5
3.0
2.5
2.0
Hole 9
860
880
900
920
940
960
980
1000
1020
player_rating

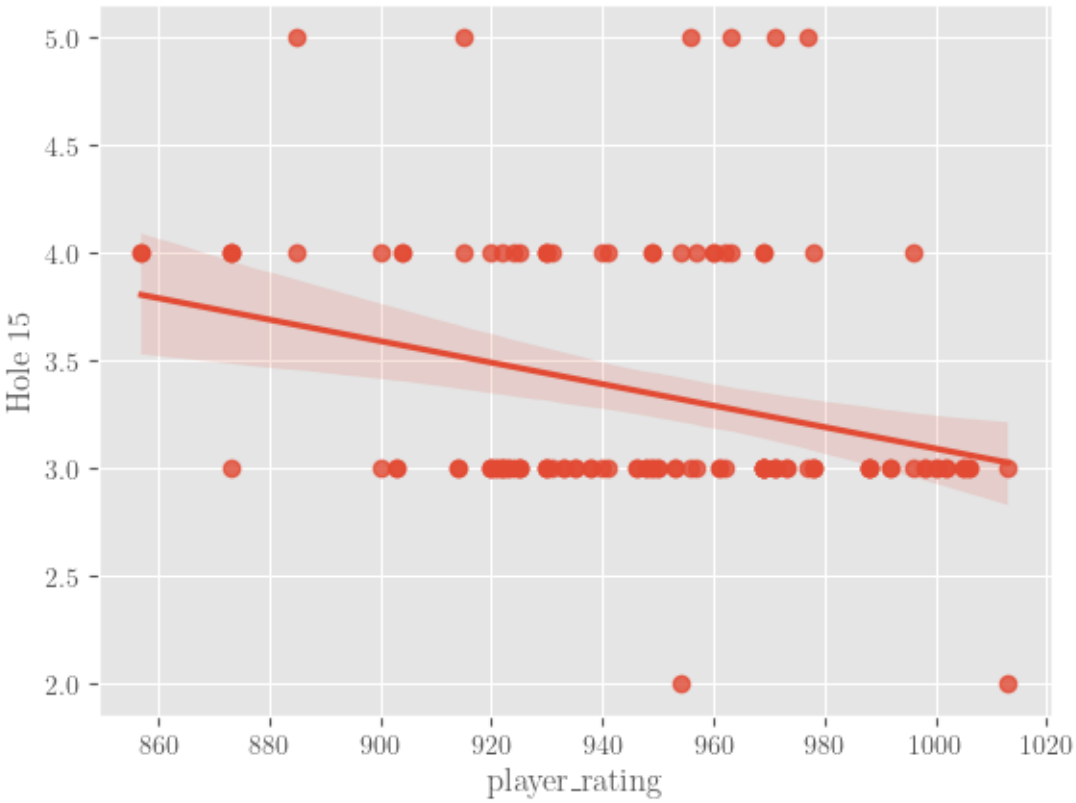

8
7
6
5
4
3
Hole 10
860
880
900
920
940
960
980
1000
1020
player_rating

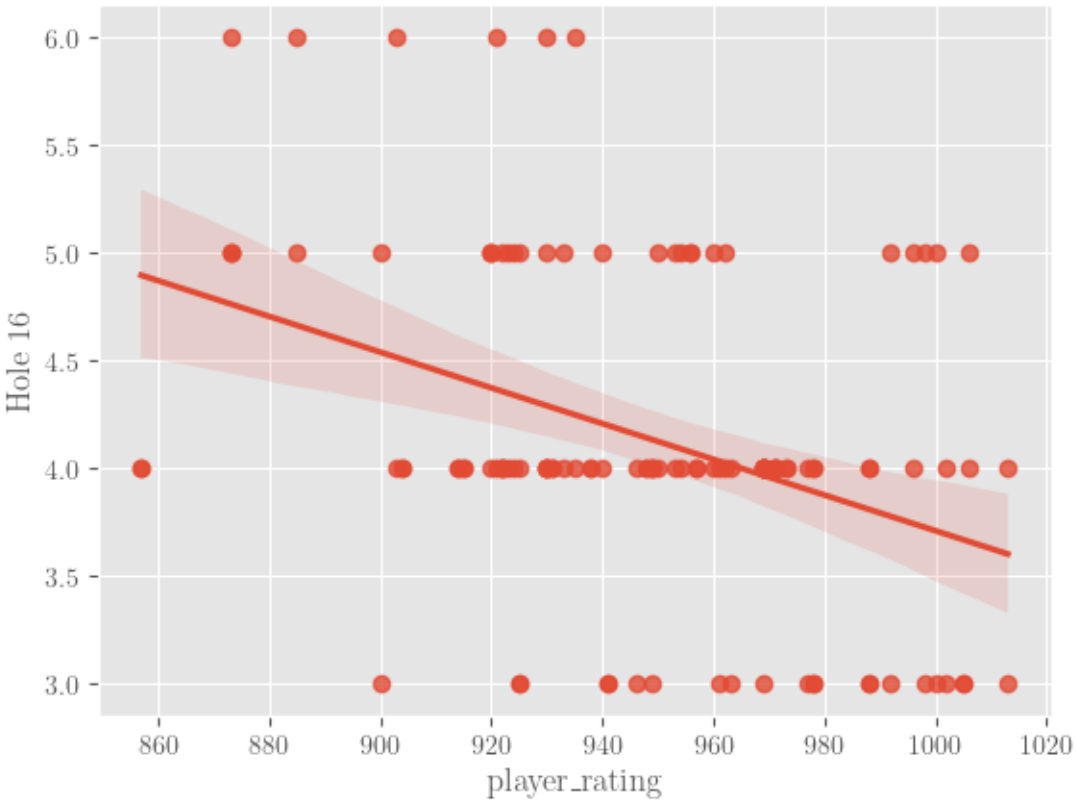

6.0
5.5
5.0
4.5
4.0
3.5
3.0
Hole 11
860
880
900
920
940
960
980
1000
1020
player_rating

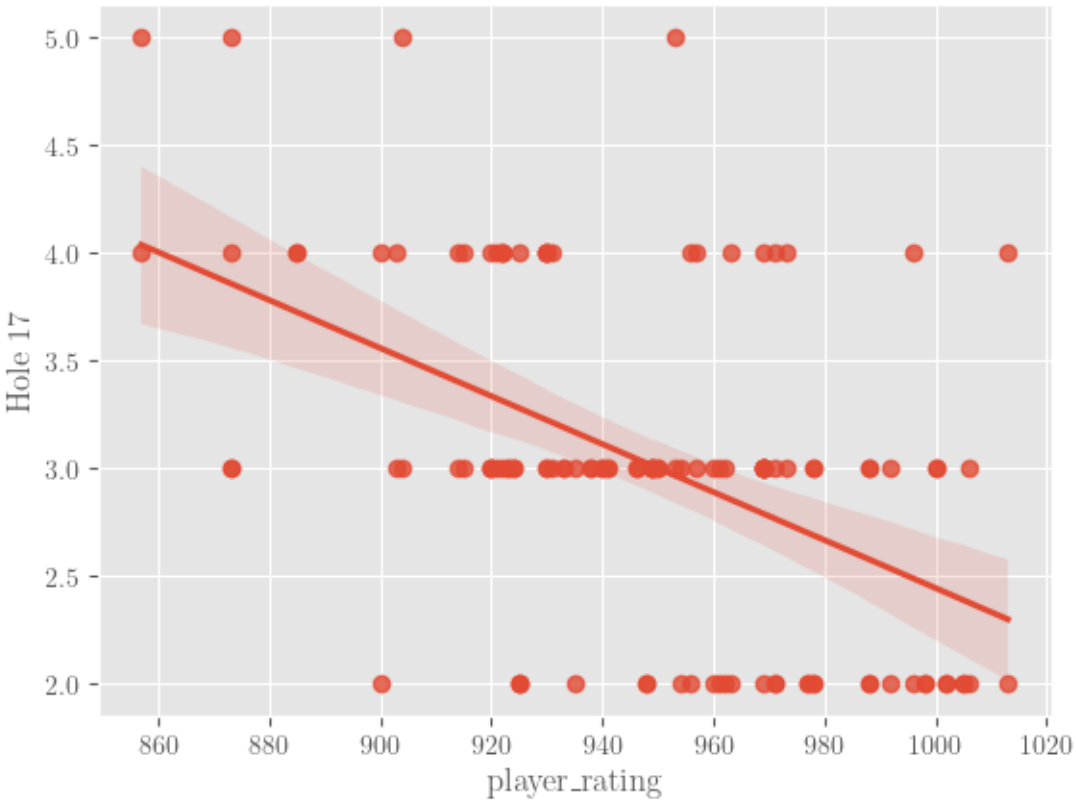

Hole 12
5.0
4.5
4.0
3.5
3.0
2.5
2.0
860
880
900
920
940
960
980
1000
1020
player_rating

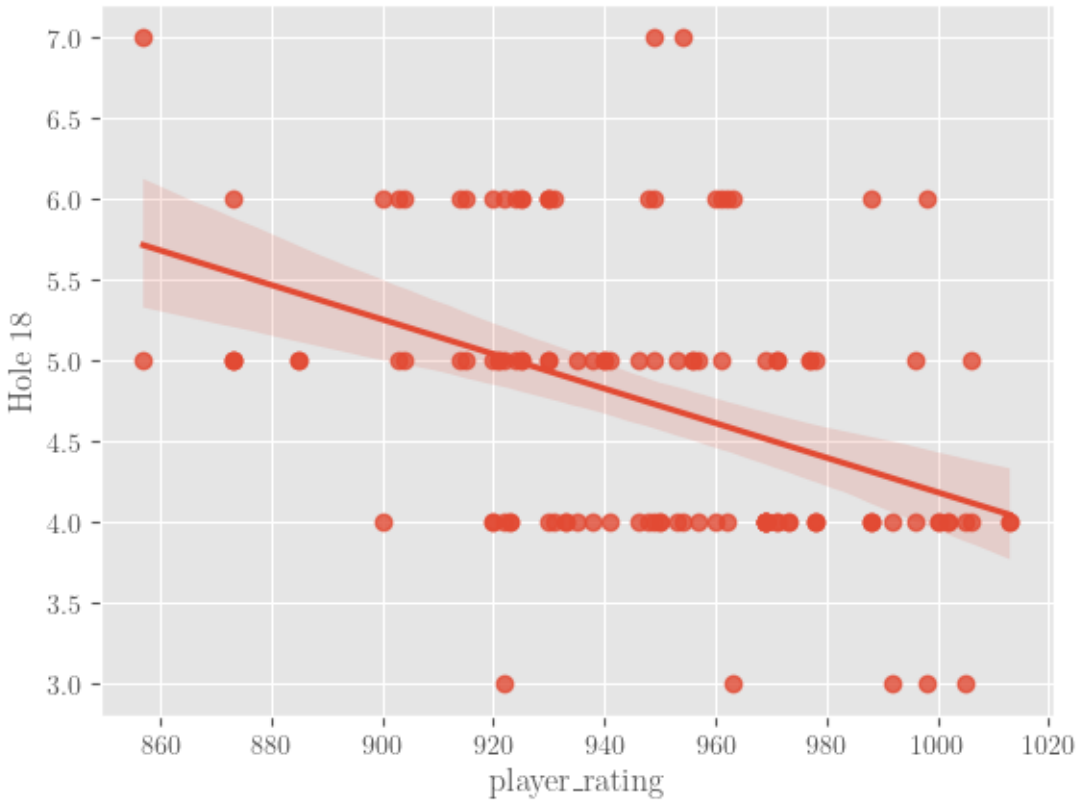

6.0
5.5
5.0
4.5
4.0
3.5
3.0
Hole 13
860
880
900
920
940
960
980
1000
1020
player_rating


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
Hole 14
860
880
900
920
940
960
980
1000
1020
player_rating

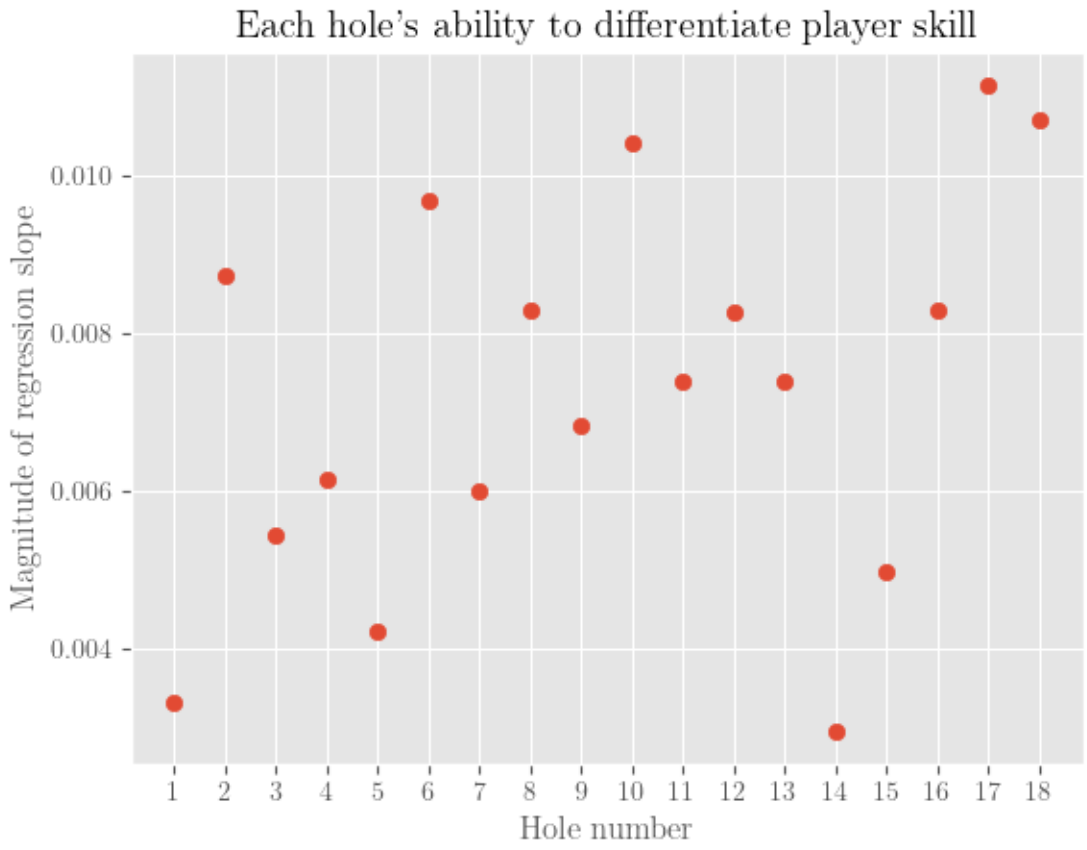

5.0
4.5
4.0
3.5
3.0
2.5
2.0
Hole 15
860
880
900
920
940
960
980
1000
1020
player_rating


6.0
5.5
5.0
4.5
4.0
3.5
3.0
Hole 16
860
880
900
920
940
960
980
1000
1020
player_rating


Hole 17
player_rating
5.0
4.5
4.0
3.5
3.0
2.5
2.0
860
880
900
920
940
960
980
1000
1020



[9]:
```
plt.scatter(slopes.keys(), [-1 * x for x in slopes.values()])
plt.xticks(range(1, 19))
plt.xlabel("Hole number")
plt.ylabel("Magnitude of regression slope")
plt.title("Each hole's ability to differentiate player skill")
plt.show()
```



[10]:
```
s = {k: round(v, 4) for k, v in sorted(slopes.items(), key=lambda item:␣
↪item[1])}
s.keys()
```

[10]:
```
dict_keys([17, 18, 10, 6, 2, 16, 8, 12, 11, 13, 9, 4, 7, 3, 15, 5, 1, 14])
```

0.2 Metadata

[11]:
```
metadata['tournament_info']
```

[11]:
```
{'DateRange': 'Sunday, Sep 14, 2025',
 'EndDate': '2025-09-14',
 'Country': 'United States',
 'Location': 'Pittstown, New Jersey',
 'LocationShort': 'Pittstown, NJ',
 'Divisions': [{'DivisionID': 100,
   'Division': 'MPO',
   'DivisionName': 'Mixed Pro Open',
   'Players': 47,
   'IsPro': True,
```

```
  'ShortName': 'Mixed Pro Open',
  'AbbreviatedName': 'Mixed Pro Open',
  'LatestRound': 2},
 {'DivisionID': 105,
  'Division': 'FPO',
  'DivisionName': "Women's Pro Open",
  'Players': 1,
  'IsPro': True,
  'ShortName': "Women's Pro Open",
  'AbbreviatedName': "Women's Pro Open",
  'LatestRound': 2},
 {'DivisionID': 121,
  'Division': 'MP50',
  'DivisionName': 'Mixed Pro 50+',
  'Players': 12,
  'IsPro': True,
  'ShortName': 'Mixed Pro 50+',
  'AbbreviatedName': 'Mixed Pro 50+',
  'LatestRound': 2}],
'Finals': 'no',
'HighestCompletedRound': 2,
'LatestRound': 2,
'FinalRound': 2,
'Layouts': [{'LayoutID': 729830,
  'CourseID': 206491,
  'CourseName': 'Camp Tecumseh',
  'Name': 'Tecumseh Takeover 2025 Pro Day Layout',
  'Holes': 18,
  'Par': 62,
  'Length': 8766,
  'Units': 'Feet',
  'Accuracy': 'M',
  'Details': [{'Hole': 'H1', 'Label': '1', 'Par': 3, 'Length': 328},
   {'Hole': 'H2', 'Label': '2', 'Par': 4, 'Length': 568},
   {'Hole': 'H3', 'Label': '3', 'Par': 3, 'Length': 371},
   {'Hole': 'H4', 'Label': '4S', 'Par': 4, 'Length': 656},
   {'Hole': 'H5', 'Label': '6', 'Par': 3, 'Length': 234},
   {'Hole': 'H6', 'Label': '7', 'Par': 5, 'Length': 755},
   {'Hole': 'H7', 'Label': '8', 'Par': 3, 'Length': 424},
   {'Hole': 'H8', 'Label': '9', 'Par': 4, 'Length': 641},
   {'Hole': 'H9', 'Label': '10', 'Par': 3, 'Length': 317},
   {'Hole': 'H10', 'Label': '11', 'Par': 4, 'Length': 503},
   {'Hole': 'H11', 'Label': '12', 'Par': 3, 'Length': 401},
   {'Hole': 'H12', 'Label': '13', 'Par': 3, 'Length': 282},
   {'Hole': 'H13', 'Label': '14', 'Par': 3, 'Length': 531},
   {'Hole': 'H14', 'Label': '16', 'Par': 3, 'Length': 336},
   {'Hole': 'H15', 'Label': 'B', 'Par': 3, 'Length': 437},
```

```
    {'Hole': 'H16', 'Label': 'C', 'Par': 4, 'Length': 639},
    {'Hole': 'H17', 'Label': '17', 'Par': 3, 'Length': 441},
    {'Hole': 'H18', 'Label': '18', 'Par': 4, 'Length': 902}]},
  {'LayoutID': 729831,
   'CourseID': 206491,
   'CourseName': 'Camp Tecumseh',
   'Name': 'Tecumseh Takeover 2025 Pro Day FPO Layout',
   'Holes': 18,
   'Par': 66,
   'Length': 8667,
   'Units': 'Feet',
   'Accuracy': 'M',
   'Details': [{'Hole': 'H1', 'Label': '1', 'Par': 3, 'Length': 229},
    {'Hole': 'H2', 'Label': '2', 'Par': 4, 'Length': 568},
    {'Hole': 'H3', 'Label': '3', 'Par': 3, 'Length': 371},
    {'Hole': 'H4', 'Label': '4S', 'Par': 4, 'Length': 656},
    {'Hole': 'H5', 'Label': '6', 'Par': 3, 'Length': 234},
    {'Hole': 'H6', 'Label': '7', 'Par': 5, 'Length': 755},
    {'Hole': 'H7', 'Label': '8', 'Par': 3, 'Length': 424},
    {'Hole': 'H8', 'Label': '9', 'Par': 5, 'Length': 641},
    {'Hole': 'H9', 'Label': '10', 'Par': 3, 'Length': 317},
    {'Hole': 'H10', 'Label': '11', 'Par': 4, 'Length': 503},
    {'Hole': 'H11', 'Label': '12', 'Par': 4, 'Length': 401},
    {'Hole': 'H12', 'Label': '13', 'Par': 3, 'Length': 282},
    {'Hole': 'H13', 'Label': '14', 'Par': 4, 'Length': 531},
    {'Hole': 'H14', 'Label': '16', 'Par': 3, 'Length': 336},
    {'Hole': 'H15', 'Label': 'B', 'Par': 3, 'Length': 437},
    {'Hole': 'H16', 'Label': 'C', 'Par': 4, 'Length': 639},
    {'Hole': 'H17', 'Label': '17', 'Par': 3, 'Length': 441},
    {'Hole': 'H18', 'Label': '18', 'Par': 5, 'Length': 902}]}],
 'Name': 'Tecumseh Takeover 2025 - Pro Day',
 'NameHtml': 'Tecumseh Takeover 2025 - Pro Day',
 'MultiLineName': {'pre': '',
  'main': 'Tecumseh Takeover 2025 - Pro Day',
  'post': ''},
 'MultiLineNameHtml': ' <span class="event-name-main">Tecumseh Takeover 2025 -
Pro Day</span>',
 'RawTier': 'B',
 'Rounds': 2,
 'RoundsList': {'1': {'Number': 1,
   'Label': 'Round 1',
   'LabelAbbreviated': 'Rd 1',
   'Date': '2025-09-14',
   'DateAbbreviated': 'Sep 14',
   'ShowDate': False},
  '2': {'Number': 2,
   'Label': 'Round 2',
```

```
  'LabelAbbreviated': 'Rd 2',
  'Date': '2025-09-14',
  'DateAbbreviated': 'Sep 14',
  'ShowDate': False}},
'Semis': 'no',
'SimpleName': 'Tecumseh Takeover 2025 - Pro Day',
'StartDate': '2025-09-14',
'TDName': 'Curvin Stine',
'TDPDGANum': 41869,
'Tier': 'B',
'FormattedTier': 'B-Tier',
'FormattedLongTier': 'B-Tier',
'TierPro': 'B',
'TierAm': 'B',
'TierX': False,
'TotalPlayers': 60,
'TournamentId': '95560',
'International': False,
'FullStats': True,
'TimeZone': 'America/New_York',
'AdditionalEventInfo': None,
'EventVideo': None,
'ScoringFormat': 'S'}
```

[ ]: