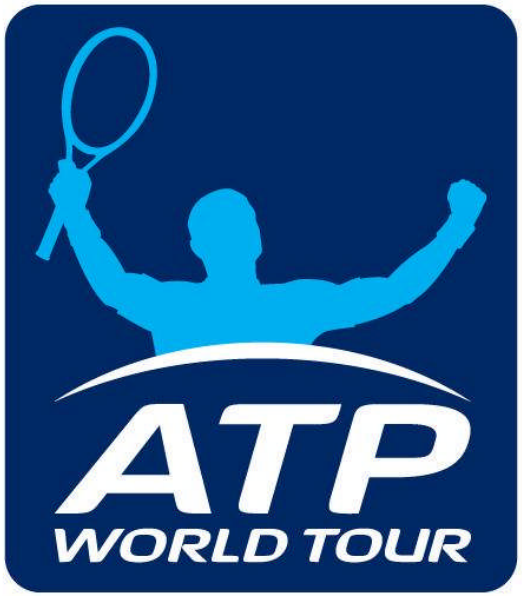
ATP 2012 – 2014 Upset Visualization

July 17, 2016



Agenda

- Context and Overview
- Odds Data Visualization
- Charted Match Statistics Visualization
- Next Steps

Agenda

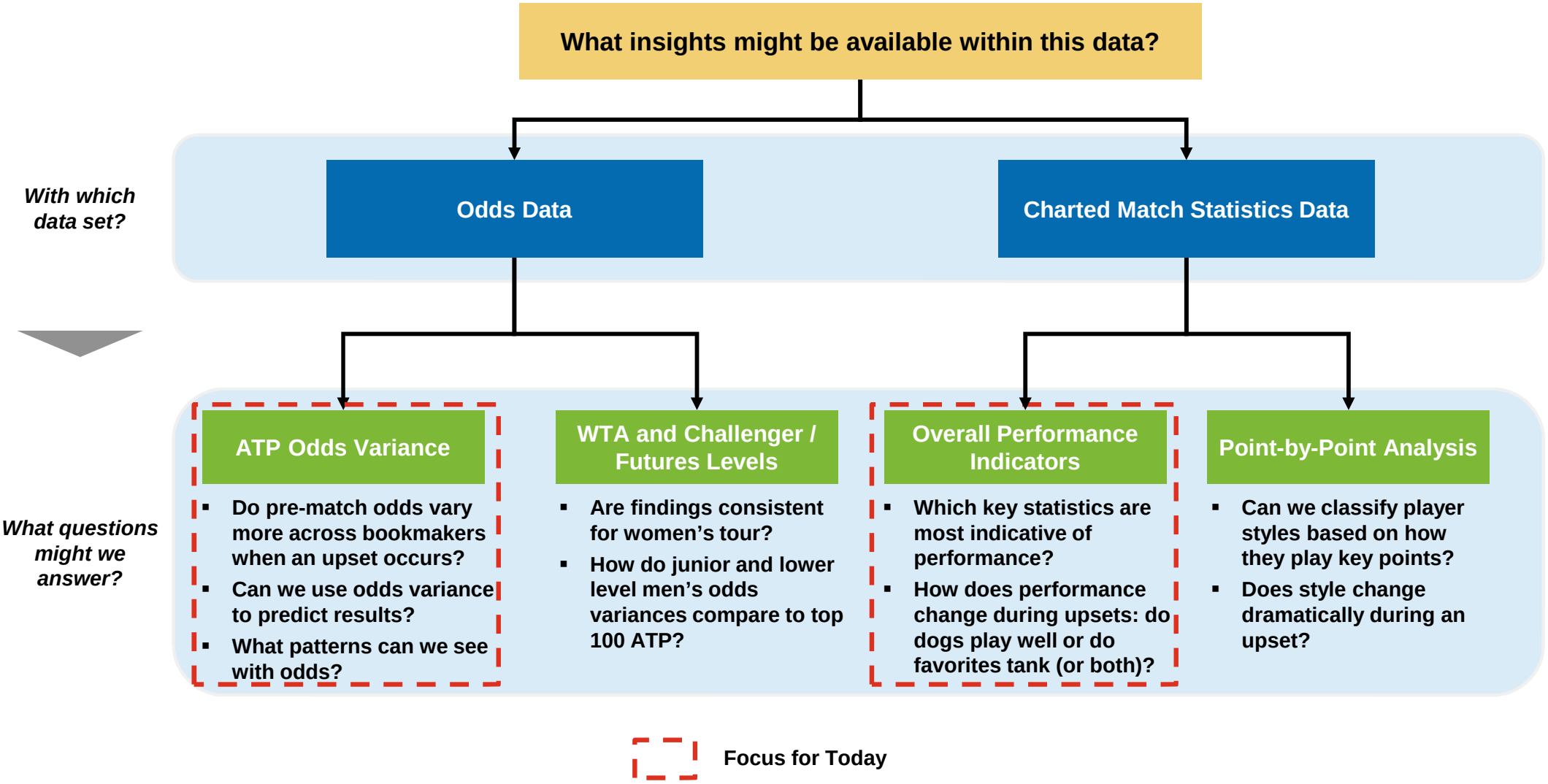
- **Context and Overview**
- Odds Data Visualization
- Charted Match Statistics Visualization
- Next Steps

Professional tennis is unique in that even despite the dominance of a few select competitors, major upsets do occur regularly

Analysis Context

- Relevant data sets are made available that represent two major categories of tennis match analysis:
  - **Odds**: published at www.tennis-data.co.uk, displays detailed match results and pre-match odds given by five major bookmakers on the day of each match
  - **Charted Statistics**: published by Jeff Sackmann at github.com/JeffSackmann/tennis_atp, a number of csv files are hosted with extremely detailed, point-by-point data for select matches
- I merged these disparate data sets for the highest level of professional men's tennis only (i.e. ATP not including challenger or future's matches) over the 2012, 2013, and 2014 seasons
- An upset is defined in this analysis as a completed match (i.e. player did not retire or walkover) in which the player with higher average odds defeated his opponent
- Visualization of this data intends to examine:
  - Whether any incongruities exist amongst odds makers that might predict when upsets are likely to occur
  - Frequency and severity of upsets
  - Performance by favorites and underdogs across key metrics during average, upset, and non-upset matches

There are a number of interesting questions that might be addressed with these data sets: today we will focus on two areas



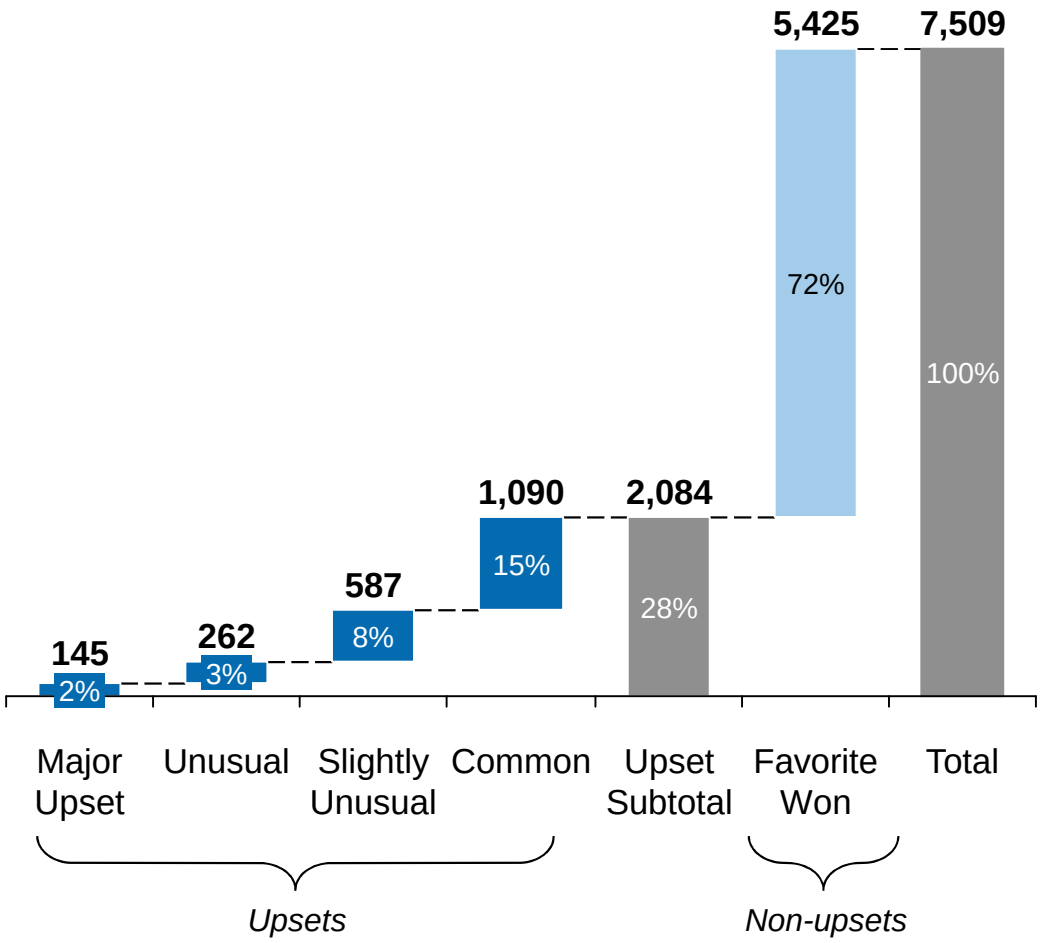
Agenda

- Context and Overview
- **Odds Data Visualization**
- Charted Match Statistics Visualization
- Next Steps

Player odds help us understand who is truly expected to win each match; favorites tend to win approximately 72% of the time

Match Waterfall: Upset vs Non-upset Frequency

(ATP Matches from 2012-2014, n = 7,509)



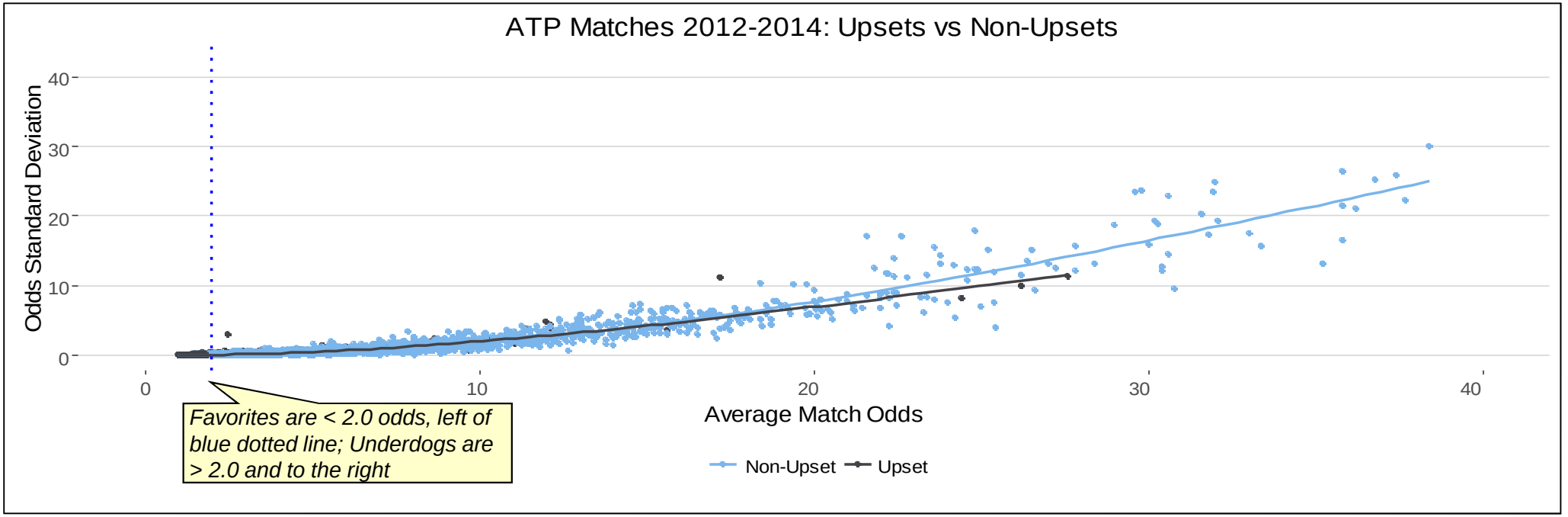
Discussion

- According to odds data, favorites win ~72% of their matches
- Remaining 28% represents underdog victories, with only 5% as either unusual or a major upset

Odds Background / Methodology

- Odds are represented in British format with favorites defined as < 2.0 and underdogs > 2.0
- This means if you were to bet $100 on a favorite with 1.9 odds and won, you would receive $190 from the bookmaker (original $100 plus $90 profit)
- Consequently, the same $100 bet on a 3.0 underdog who won would pay $300 (including $200 profit)
- Assignment of players into groups is based on their average odds:
  - **Favorites**: “Slight Favorite” < 2.0, “Clear Favorite” < 1.85, “Heavy Favorite” < 1.4, and “Sure Thing” < 1.1
  - **Underdogs**: “Slight Dog” > 2.0, “Clear Dog” > 2.6, “Heavy Dog” > 3.5, “No Hope” > 5.0

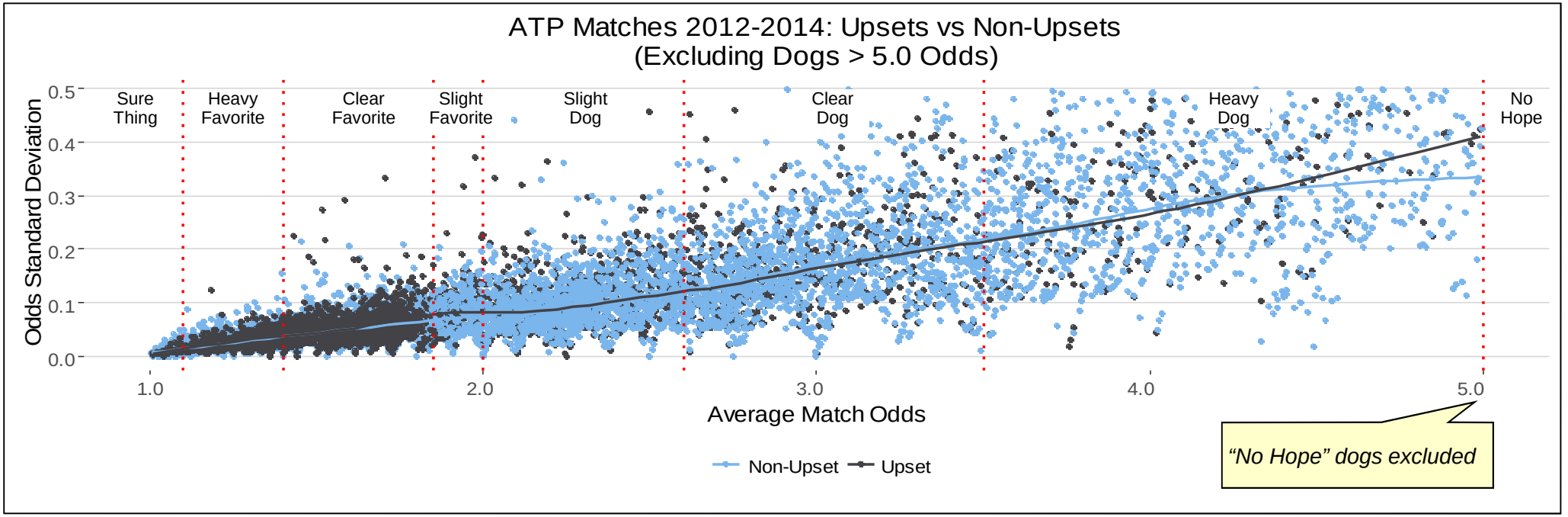
There is a clear relationship between average odds and standard deviation of odds: the bigger the underdog, the more odds vary



Discussion

- As expected, the majority of upsets (black dots) occur near the 2.0 mark, meaning the quality of players is relatively close; though there are some cases where underdogs in the 20:1 range defeated their opponents!
- In general, it seems that the amount of variance among bookmakers is much greater for major underdogs that have relatively little hope of beating the much higher ranked player
- At this level it is difficult to ascertain whether there is a meaningful difference in typical variance in upsets vs non-upsets. At first glance, it appears that the amount of variance is in line for both cases as both trend lines seem to overlap

Excluding dogs with odds > 5.0, we can drill down and confirm that odds variance is similar for both upsets and non-upsets



Discussion

- Within the 1.0 to 5.0 odds range, we again see much "tighter" odds for favorites across bookmakers, especially < 1.5 odds
- This could suggest that casinos put heavier focus on correctly handicapping higher probability events; that is, given limited time and resources, they may assume that odds accuracy for underdogs is less important since they are less likely to win
- While this data indicates we cannot predict upsets based on odds variance alone, it is interesting to note just how much variance exists for underdogs; bettors may use this information to carefully select bookmakers and maximize value when betting on underdogs

Agenda

- Context and Overview
- Odds Data Visualization
- **Charted Match Statistics Visualization**
- Next Steps

Using the match charted data available we can track player performance against 9 key metrics

Aces

- Serves that are winners and untouched by opponent
- Typically a function of both pace (speed) and placement

Winners vs Unforced Errors

- Common statistic referenced in every match
- Winner = winning point untouched by opponent
- Unforced Error = losing point in a situation where shot should have been made

Double Faults

- Considered an unforced error, free point given to opponent when neither first nor second serve is in
- Low frequency rate at pro level

Break Points Faced

- # of times a player faced a potential losing point on his own serve
- Generally difficult to "break" serves of highest caliber players

First Serves In

- Percentage of first serves that are not faults
- First serves are typically more aggressive than second serves

Break Points Won

- # of times a returner successfully "broke" his opponent's serve (won the game as a returner)

First Serves Won

- Percentage of points that are won off first serves
- High rate of this metric indicates strong server (pace / direction)

Return Points Won

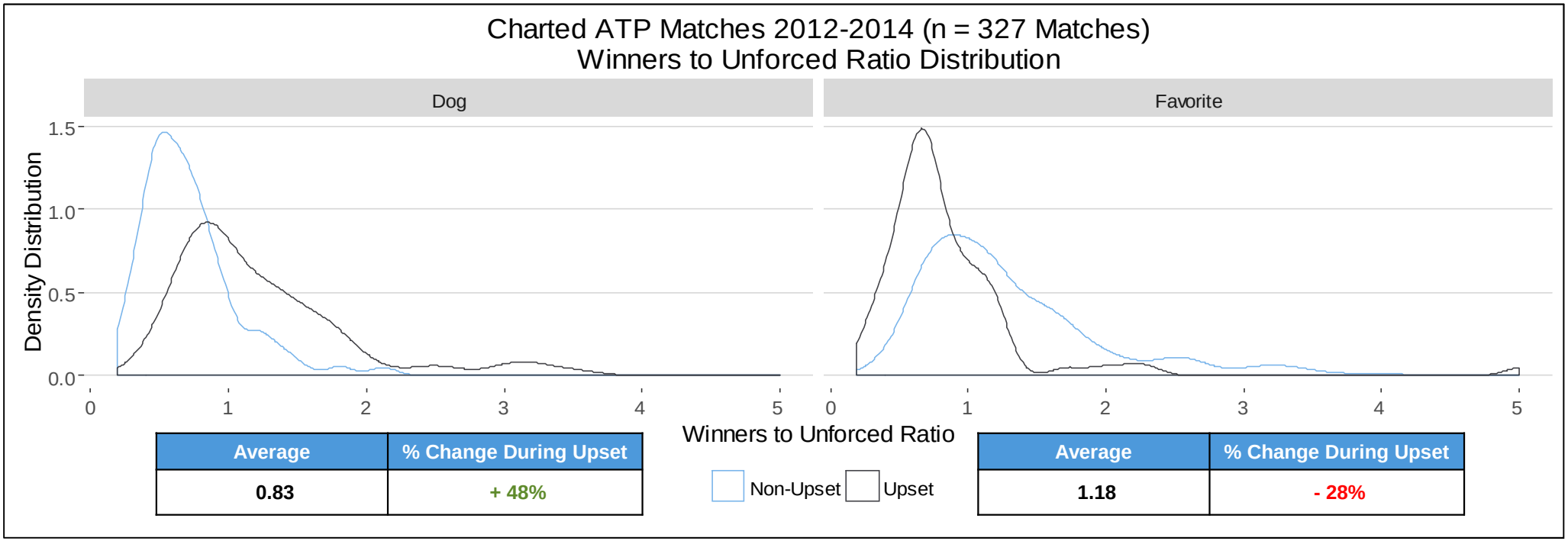
- Total number of points a returner won against his opponent's serve

Second Serves Won

- Percentage of points that are won off second serves

Highest % Change During Upset

Winners to Unforced Errors ratio is the most cited statistic in tennis for good reason: it is highly correlated to match results



Average	% Change During Upset
0.83	+ 48%

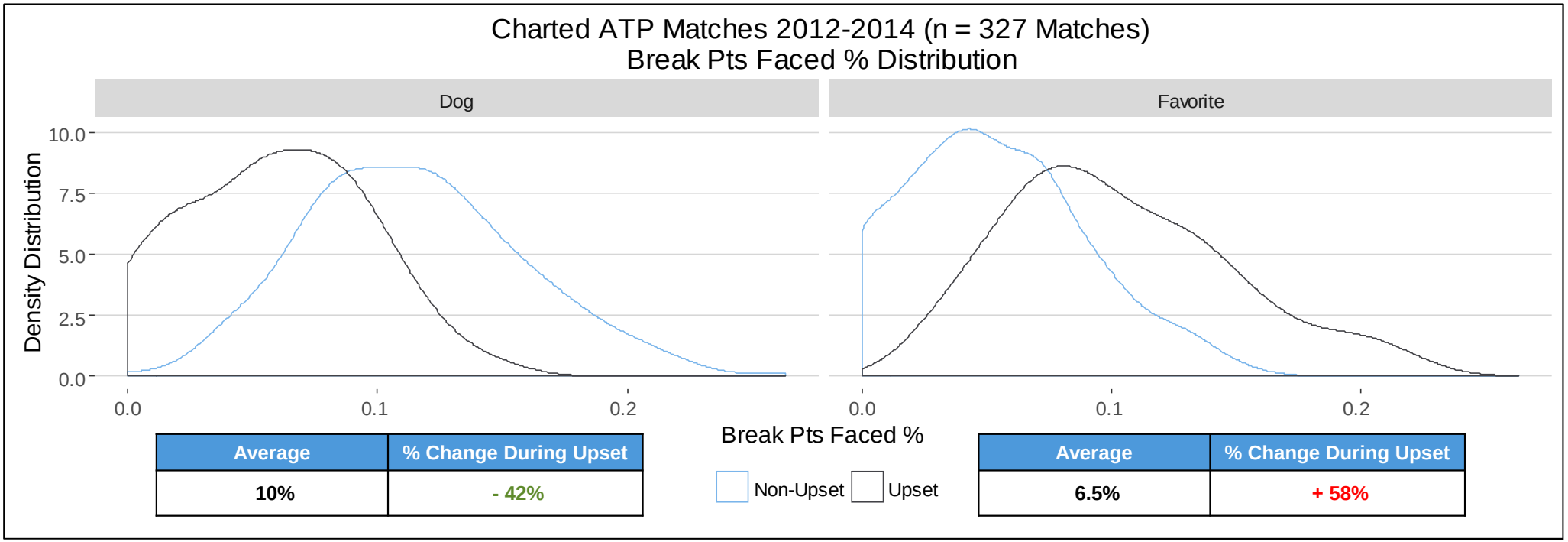
Average	% Change During Upset
1.18	- 28%

Discussion

- Winners and unforced errors constitute roughly 70% of total points played[1] and thus logically have bearing on match outcomes
- Underdogs average an unfavorable 0.83 ratio typically, though tend to improve by 48% during upsets
- This statistic would support a hypothesis that underdog performance increases more than favorite performance decreases during upsets; i.e., that the matches weren't just "bad days" for the higher ranked players

1) "A closer look at the winner-unforced error ratio" tennisabstract.com

Interestingly, break points faced is a more telling upset statistic than break points saved



Average	% Change During Upset
10%	- 42%

Average	% Change During Upset
6.5%	+ 58%

Discussion

- In this case, both favorites and underdogs see dramatic performance changes during upsets, specifically to their serves
- Underdogs tend to face 42% fewer break points while favorites see 58% more
- This underscores the fact that a strong service game is critical at the pro level and can make a world of difference
- It is also generally understood that service games are more tiring for the serving player, so fewer break points faced suggests more efficient service games, which could help the underdog maintain stamina

Agenda

- Context and Overview
- Odds Data Visualization
- Charted Match Statistics Visualization
- **Next Steps**

A fount of interesting data remains to be analyzed with time permitting

Next Steps

- Additional analyses / hypotheses that could be tested include:
  - Women's / challengers / futures tours (repeat existing visualizations)
  - Bookmaker-specific analysis (e.g., do certain books predict upsets more frequently?)
  - Point-by-Point analysis (e.g., has underdog recently changed tactics that could help us predict victory?)
  - Prior year analysis (data exists back to 1968)
  - Many more

- Questions?