Assignment 3 Example Report: Exploratory Analysis of OpenPowerlifting

Introduction

This report explores the OpenPowerlifting database, a public-domain archive of powerlifting competition results. The version used here (downloaded October 2026, with results through 13 October 2026) has 4,050,248 entries from about 1.02 million lifters. Each row is one lifter at one meet, with sex, age, body weight, the equipment category, the best successful squat, bench press and deadlift, the total of the three, and the date and federation of the meet. The data are compiled by volunteers from federation results, so coverage depends on which federations have been entered, and many fields (especially age) are incomplete.

K
Oct 18, 9:12 AM

Data Quality — Excellent

The introduction names the dataset's limits up front: coverage depends on which federations are entered, and many ages are missing. This sets up the checks later in the report; keep doing this.

Initial Analysis Questions

1. **Overarching Question:** Who competes in powerlifting, and what factors shape how much a lifter can lift?
2. How has participation in powerlifting changed over time, for men and for women?
3. How does strength change with age, and when do lifters peak?
4. How much does body size explain strength, and do scoring formulas like Dots remove that effect?
5. How much does supportive equipment (knee wraps, single-ply and multi-ply suits) add to what a lifter can total?

Discoveries & Insights

I started with an overview of the whole dataset before focusing on raw (unequipped) full-power meets, where lifters compete in all three lifts.

K
Oct 18, 9:20 AM

Breadth of Exploration — Excellent

You explain why the analysis narrows to raw full-power meets. Also state how many entries that leaves (about 1.1 million) so readers know the scope of the charts that follow.

1. Participation has grown more than sixfold since 2009

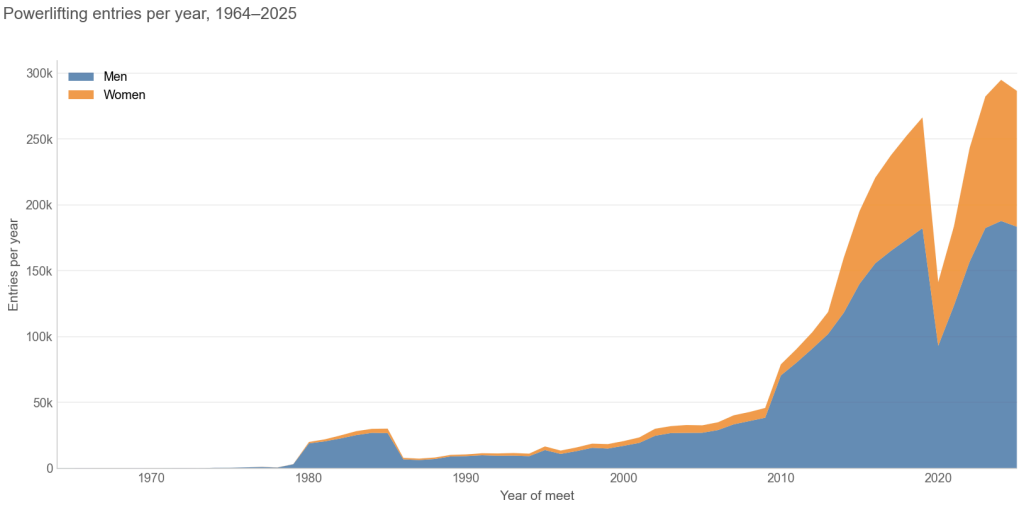


Figure 1. Number of entries per year by sex, all events and equipment, 1964–2025 (2026 omitted as a partial year). Entries grew more than sixfold between 2009 and 2025, apart from a sharp drop in 2020.

Entries per year were below 50,000 until 2009, then rose steeply to 286,629 in 2025. I excluded 2026 because the data stop in October. Two drops stand out: entries fell 47% in 2020, when meets were cancelled during the COVID-19 pandemic, and they fell from 30,165 in 1985 to 7,988 in 1986, almost all within the USPF federation, which looks more like a gap in the archive than a real drop in participation. Women made up 5% of entries in 1980 and 36% in 2025.

2. Most entries are full-power meets, but equipment use is mixed

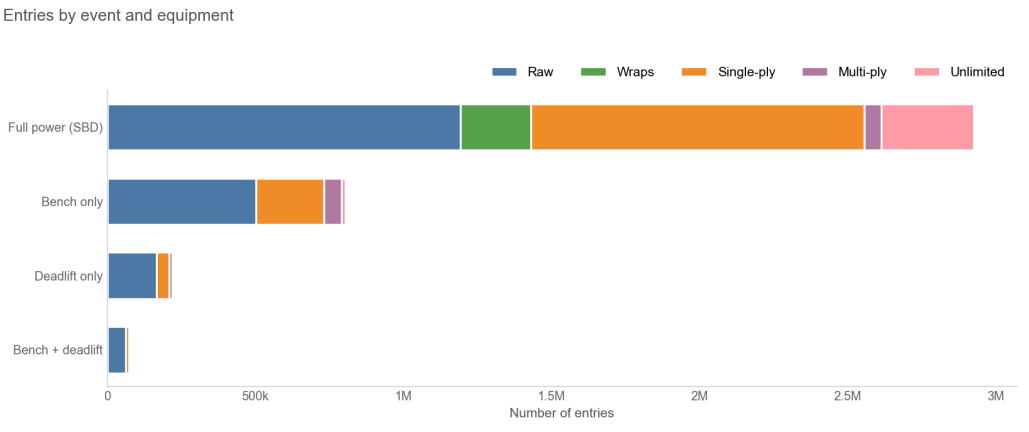


Figure 2. Entries by event and equipment.

K
Oct 18, 9:34 AM

Data Quality — Excellent

Tracing the 1985–1986 drop to a single federation separates a gap in the archive from a real change in participation. This is the kind of sanity check the rubric means by thorough profiling.

K
Oct 18, 2:05 PM

Captions — Satisfactory

This caption only repeats the chart title. A caption should say what the chart shows and what to notice, for example that full-power meets are 72% of entries and that raw and single-ply are their two largest equipment types. Add one sentence with the main point.

Full-power meets (squat, bench press and deadlift) account for 72% of entries, followed by bench-only meets. Within full power, raw lifting is the largest category, but single-ply is close behind. Because equipment changes how much people can lift, I kept the categories separate for the rest of the analysis.

3. Age is missing or approximate for most entries

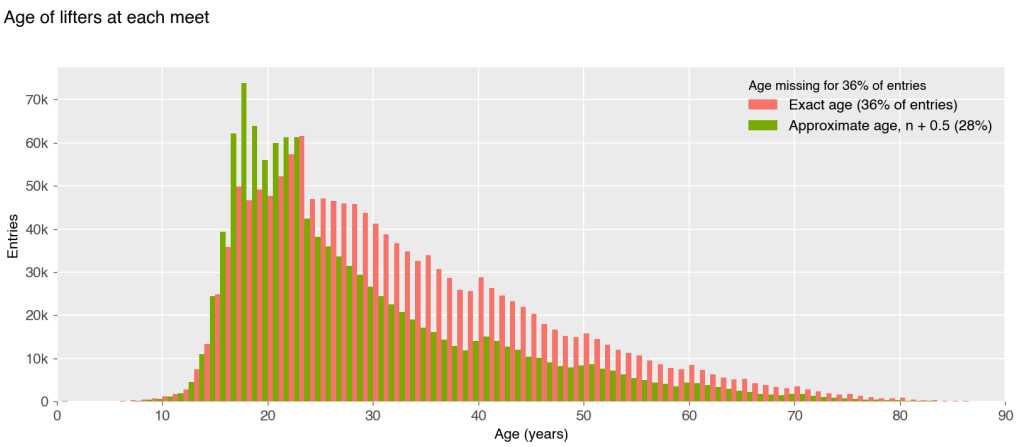


Figure 3. Lifter age at each meet in 0.5-year bins, split into exact ages (red) and approximate ages ending in .5 (green). Approximate ages are concentrated among teenagers and young adults, and 36% of entries have no age at all.

Age is missing for 36% of entries. Another 28% have an "approximate" age ending in .5, which OpenPowerlifting uses when a federation only reports birth year. The green bars show that approximate ages are concentrated among teenagers and young adults, probably because school and youth federations report birth years. I also found 453 entries with an age under 5, which are almost certainly data entry errors. For the age analysis below, I kept ages 14 to 80 and rounded approximate ages.

K
Oct 18, 2:11 PM

Visualizations — Satisfactory

In a five-segment stacked bar, only the first segment (Raw) sits on a common baseline, so the other equipment types cannot be compared accurately across events. To reach Excellent, use a grouped bar chart on a zero baseline, or a 100% stacked bar if the point is each equipment type's share within an event.

K
Oct 19, 10:02 AM

Data Quality — Excellent

Separating exact from approximate ages and flagging 453 ages under 5 profiles the field before you rely on it. Report how many entries the 14–80 filter removes so the effect of cleaning is clear.

4. Men's body weights pile up just under the weight-class limits

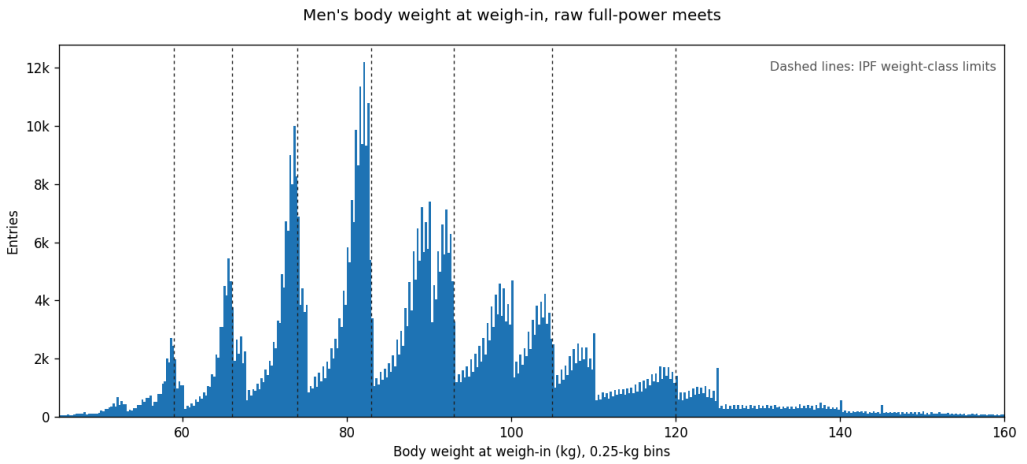


Figure 4. Body weight at weigh-in for men in raw full-power meets, in 0.25-kg bins, with dashed lines at the IPF weight-class limits. Counts spike just below each limit, showing that lifters cut weight to make a class.

Instead of a smooth curve, men's body weights form a sawtooth: counts rise sharply just under each weight-class limit (59, 66, 74, 83, 93, 105 and 120 kg) and drop right after it. Lifters cut weight to compete in the lightest class they can make. This matters for any analysis of body weight, because people are not spread evenly across the scale.

5. Heavier lifters lift more, but the gain flattens

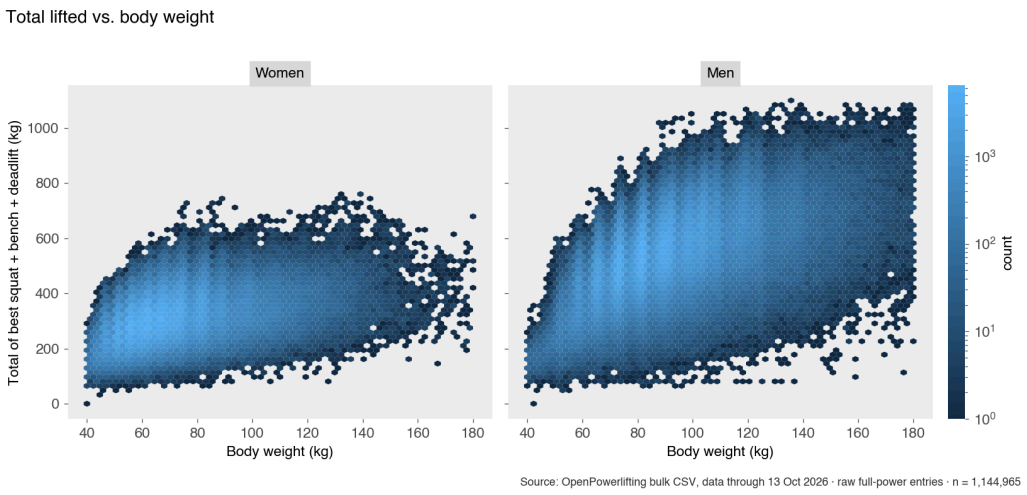


Figure 5. Total (best squat + bench press + deadlift) against body weight for raw full-power entries, by sex, shown as hexagonal bins with a log color scale. Total rises with body weight for both sexes, but the rise is steeper for men.

K
Oct 19, 10:15 AM

Visualizations — Satisfactory

The fine bins make the sawtooth visible, but the dashed reference lines are not labeled, so readers have to guess which class each one marks. Label each line with its limit (59, 66, 74 kg and so on) directly on the chart.

I used a hexbin plot with a log color scale because a scatter plot of over a million points was a solid blob. Total rises with body weight for both sexes (correlation 0.64 overall), but the increase slows above about 100 kg for men and 80 kg for women. The vertical stripes come from the weight-class spikes in the previous chart.

6. Dots mostly removes the effect of body weight for men, less so for women

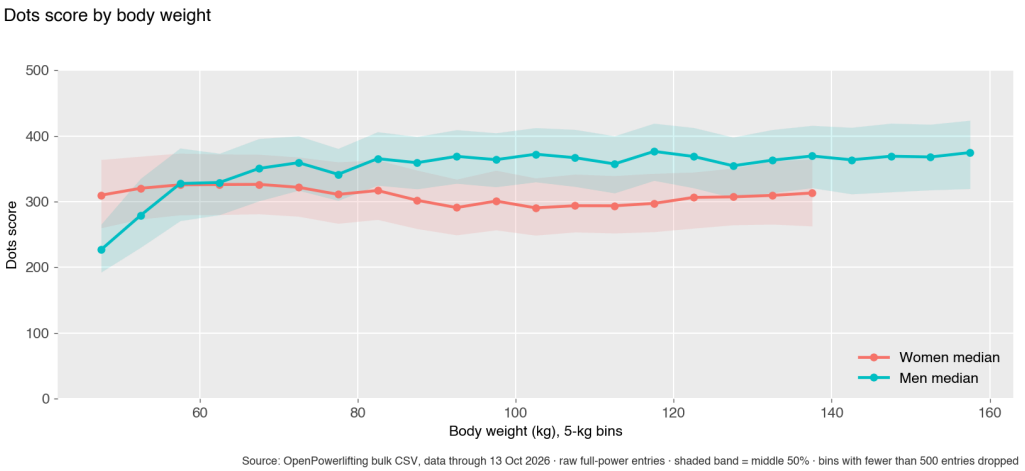


Figure 6. Median Dots score (line) and middle 50% (band) by 5-kg body-weight bin and sex, raw full-power entries. Men's Dots is nearly flat above about 75 kg but lower for the lightest men.

Dots is a formula that adjusts a total for body weight so lifters of different sizes can be compared. Its correlation with body weight is only 0.18, compared with 0.64 for the raw total. For men, median Dots is nearly flat from 75 kg upward, but it is clearly lower below 60 kg. For women, Dots is highest in the lighter classes and dips between 85 and 120 kg. So Dots works reasonably well but still favors some body sizes over others.

K
Oct 19, 10:40 AM

Visualizations — Satisfactory

Hexbins with a log color scale handle over a million points well. The caption's claim that gains slow above 100 kg cannot be seen from density alone; overlay a median line by body weight so the claim is visible in the chart.

K
Oct 19, 11:03 AM

Depth of Exploration — Excellent

Testing whether Dots actually removes the effect of body weight, instead of assuming it does, is a strong follow-up question. A next step would be to check Wilks, which is in the same file, to see whether the dip for heavier women is specific to Dots.

7. Strength peaks around 30 and then declines steadily

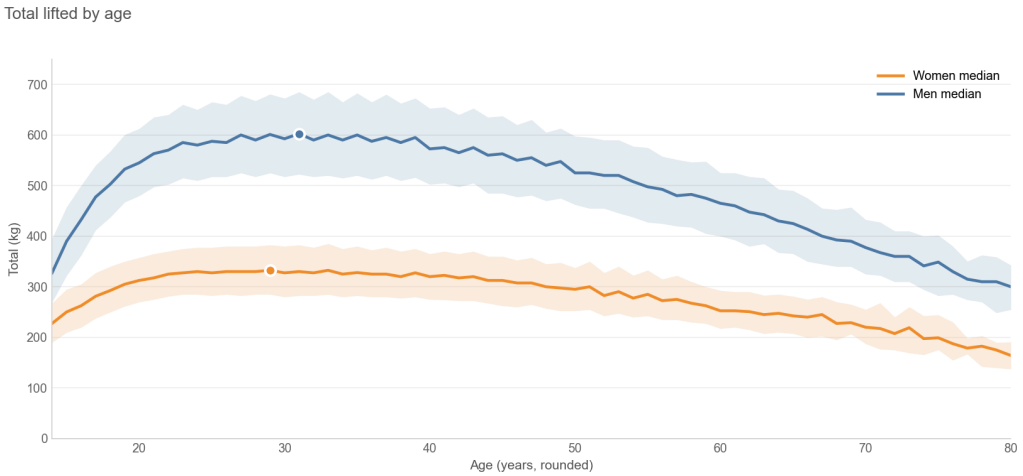


Figure 7. Median total (line) and middle 50% (band) by age for raw full-power entries with a known age; dots mark each peak. Totals peak at 31 for men and 29 for women and fall steadily after 40.

Median total rises quickly through the teens, peaks at age 31 for men (602.5 kg) and 29 for women (332.5 kg), and then declines. By age 70, median totals are about 37% lower for men and 34% lower for women than at the peak. The middle-50% band stays wide at every age, so many older lifters out-total much younger ones.

8. Women get a larger share of their total from the deadlift

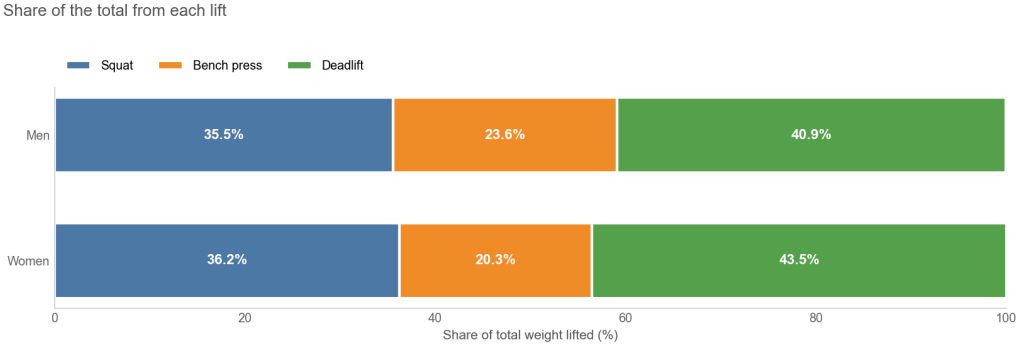


Figure 8. Squat, bench press and deadlift shares of the total.

K
Oct 19, 3:22 PM

Depth of Exploration — Excellent

The numbers are clear and well supported. Note that this compares different people at each age, not the same people getting older, so "declines" describes lower totals in older groups; point ahead to chart 11, which follows the same lifters over time.

K
Oct 19, 3:47 PM

Captions — Satisfactory

The caption does not say which entries are included (raw full-power entries, both sexes) or what readers should take from the chart. Add the data filter and the key difference, such as the women's deadlift share being about 2.6 percentage points higher.

The deadlift is the largest share of the total for both sexes, but more so for women (43.5% vs. 40.9%), while the bench press is a smaller share for women (20.3% vs. 23.6%). This fits the idea that upper-body strength differs more between the sexes than lower-body strength.

9. Surprisingly, single-ply lifters do not total more than raw lifters

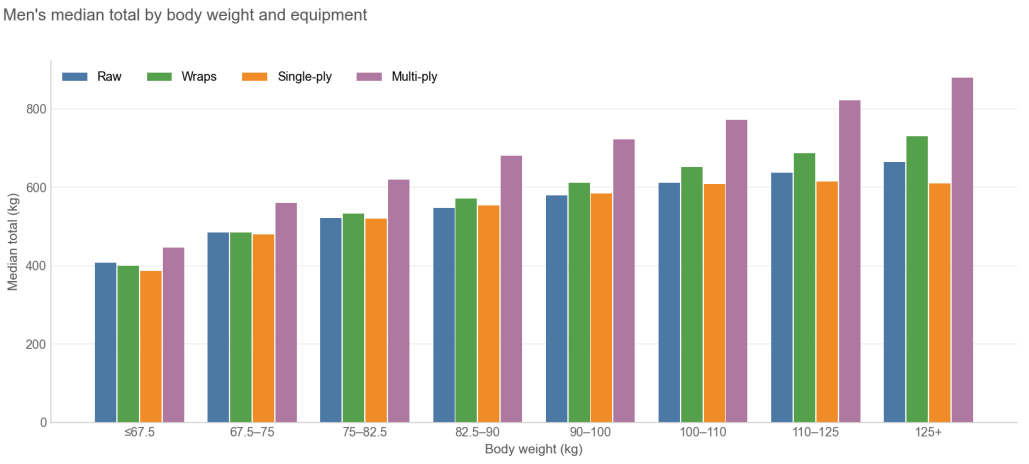


Figure 9. Median total for men by body weight and equipment.

I expected equipped lifters to total much more, since supportive suits store energy. Multi-ply lifters do total far more (for example, 822 kg vs. 638 kg at 110–125 kg), and wraps add a little. But single-ply medians are about the same as raw, and even lower in the heaviest classes. My guess is that the groups are not comparable: single-ply entries are mostly from older years (median 2013, vs. 2022 for raw) and younger lifters (median age 22.5 vs. 25).

K
Oct 20, 8:55 AM

Visualizations — Satisfactory

Two bars for all 1.1 million entries hide how lift shares vary by weight class or age. To reach Excellent, show the shares by weight class or age in small multiples. The text also attributes the difference to upper-body strength, which this chart cannot show; state it as a hypothesis to test next.

K
Oct 20, 9:10 AM

Captions — Satisfactory

The caption leaves out the context that matters most: the bars mix entries from every year, and most single-ply entries are from earlier years. Someone reading only the figure and caption would conclude that single-ply gear does not help. State the time range and that the equipment groups differ in era and age.

K
Oct 20, 9:31 AM

Depth of Exploration — Excellent

Flagging that this chart invites a wrong conclusion, and giving evidence (era and age differences), is good analysis that leads directly to the follow-up in chart 10. The text jumps from lift shares to equipment without a transition, though; one sentence linking back to the equipment question would help readers follow.

10. The same lifters total more in single-ply, but not always

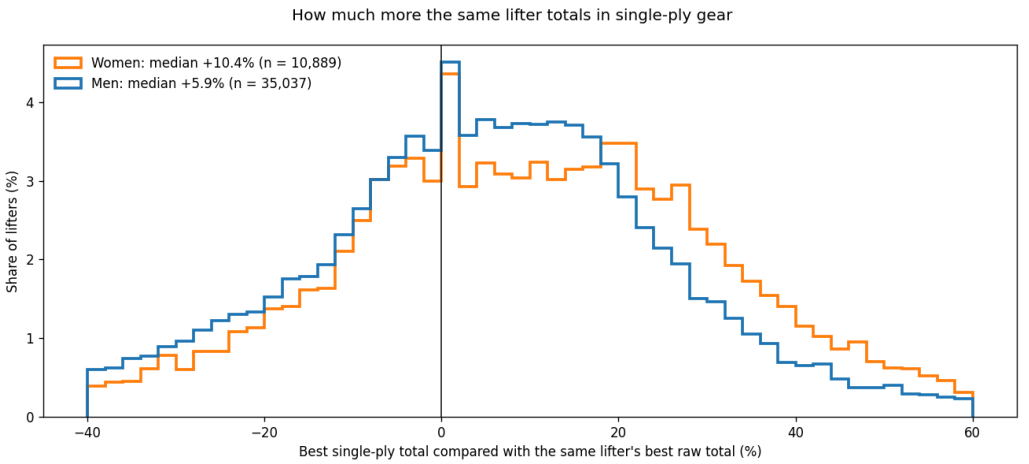


Figure 10. Each lifter's best single-ply total relative to their best raw total, for lifters with both, by sex; the 8% of lifters outside −40% to +60% are not shown. Most lifters total more in single-ply, but a large minority total less.

To remove the comparison problem in the previous chart, I matched lifters who had at least one raw and one single-ply full-power total and compared each lifter's best in each. The median lifter totals 5.9% more in single-ply (men) and 10.4% more (women). However, 37% of these lifters totaled less in single-ply. Among them, 77% did their raw meets later in their career, so they had probably gotten stronger by the time they switched to raw.

K
Oct 18, 9:12 AM

Data Transformation — Excellent

Matching each lifter's best raw and single-ply totals turns a confounded group comparison into a within-person one, and checking the order of meets explains the negative cases. Label this explicitly as a follow-up question to chart 9 so readers see how the analysis deepened.

11. Lifters improve about 20% over their first ten meets

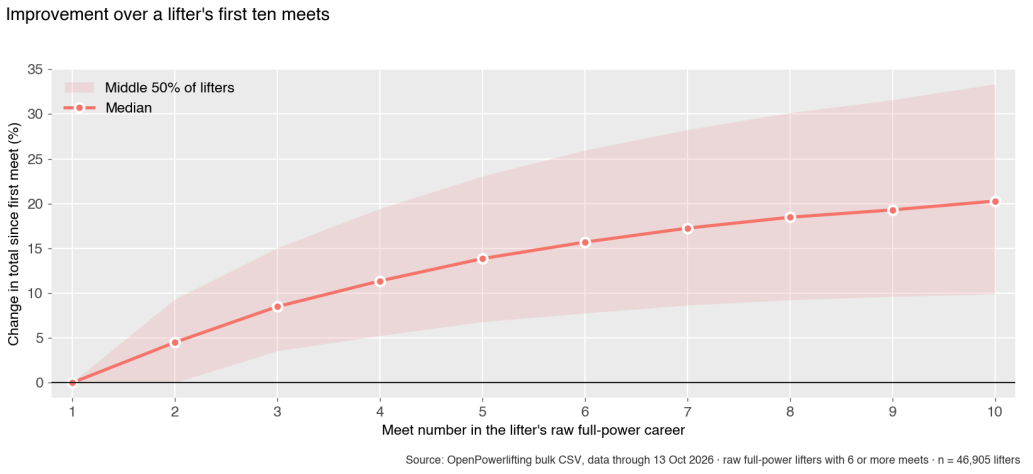


Figure 11. Median change in total since the first meet (line) and middle 50% (band) by meet number, for raw full-power lifters with six or more meets. Gains are largest over the first few meets and then level off.

For the 46,905 raw full-power lifters with at least six meets, the median total rose 4.5% by the second meet, 13.9% by the fifth and 20.3% by the tenth, with the gains slowing each time. Only lifters who keep competing are included, so this probably overstates how much a typical beginner improves.

Summary

Powerlifting has grown rapidly since 2009, especially among women, and the data reflect real patterns in strength: lifters peak around 30, heavier lifters lift more but with diminishing returns, and equipment adds to what the same person can lift. The biggest lessons were about the data themselves. Age is missing or estimated for most entries, body weights cluster at weight-class limits, coverage gaps create false drops in participation, and simple comparisons between groups (like raw vs. single-ply) can point in the wrong direction when the groups differ in age, era or federation. Comparing the same lifters over time gave more trustworthy answers than comparing groups.

Grading

Each rubric component is worth 1 point: Excellent = 1, Satisfactory = 0.5, Poor = 0. Comments name the component they affect; individual charts do not carry separate points.

Component	Score	Rating
Breadth of Exploration	1	Excellent
Depth of Exploration	1	Excellent
Data Quality	1	Excellent
Visualizations	0.5	Satisfactory
Data Transformation	1	Excellent
Captions	0.5	Satisfactory
Creativity & Originality	0.5	Satisfactory
Total	**5.5 / 7**	

K
Oct 18, 9:20 AM

Data Quality — Excellent

Naming survivorship bias shows you understand who is in this sample. Splitting by sex or starting age would show whether everyone improves at the same rate.

K
Oct 18, 9:34 AM

Creativity & Originality — Satisfactory

The summary ties the findings to lessons about the data well, but the analysis follows a standard exploratory path with standard chart types. To reach Excellent, add an original angle, such as a new derived measure (for example, each total relative to the record in its weight class) or annotations that guide the reader to the key points.

Grader summary notes

- **Strengths:** Excellent data quality profiling (missing and approximate ages, impossible ages, the 1985–1986 archive gap, weight-class spikes) and an advanced transformation that matches the same lifters across equipment types (chart 10) and across meets (chart 11).
- **Visualizations (0.5/1):** Charts 2, 4, 5 and 8 use encodings or labeling that make the stated claim hard to see: shifting baselines in a five-segment stacked bar, unlabeled reference lines, no trend line on the hexbin, and two aggregate bars that hide variation.
- **Captions (0.5/1):** Most captions describe the chart and state its main point, but the captions for Figures 2, 8 and 9 only restate the title or leave out the data filter and key context. On its own, the Figure 9 caption leads to the wrong conclusion.
- **Creativity & Originality (0.5/1):** The analysis follows a standard exploratory workflow with standard chart types; an original derived measure or annotations that guide the reader would raise it to Excellent.