Probability & Statistics Notes

Two-way data

Remember that one-way data was **variable** data given for **individuals**. For example, people and their heights, or ice cream flavors and the number of scoops of each sold.

In this section we're transitioning to talk about two-way data, which is data which is dependent on more than one thing.

Sometimes we talk about two-way data in terms of **independent variables** and **dependent variables**.

In the case of one-way data, we had one independent variable, called the individuals, and one or more dependent variables, called the variables. In the case of two-way data, we have two independent categories on which the variables are dependent.

The difference between one-way data and two-data is easiest to visualize by comparing a data table for each one.

Two-way tables

For example, let's say that an artist started a business selling his paintings at the beginning of 2014. By the end of 2017, he's been in business for four years, and his business has grown every year. If we add up the revenue earned all four years in January, all four years in February, all four years in March, etc., we could summarize his total earnings by month in a **one-way table**:
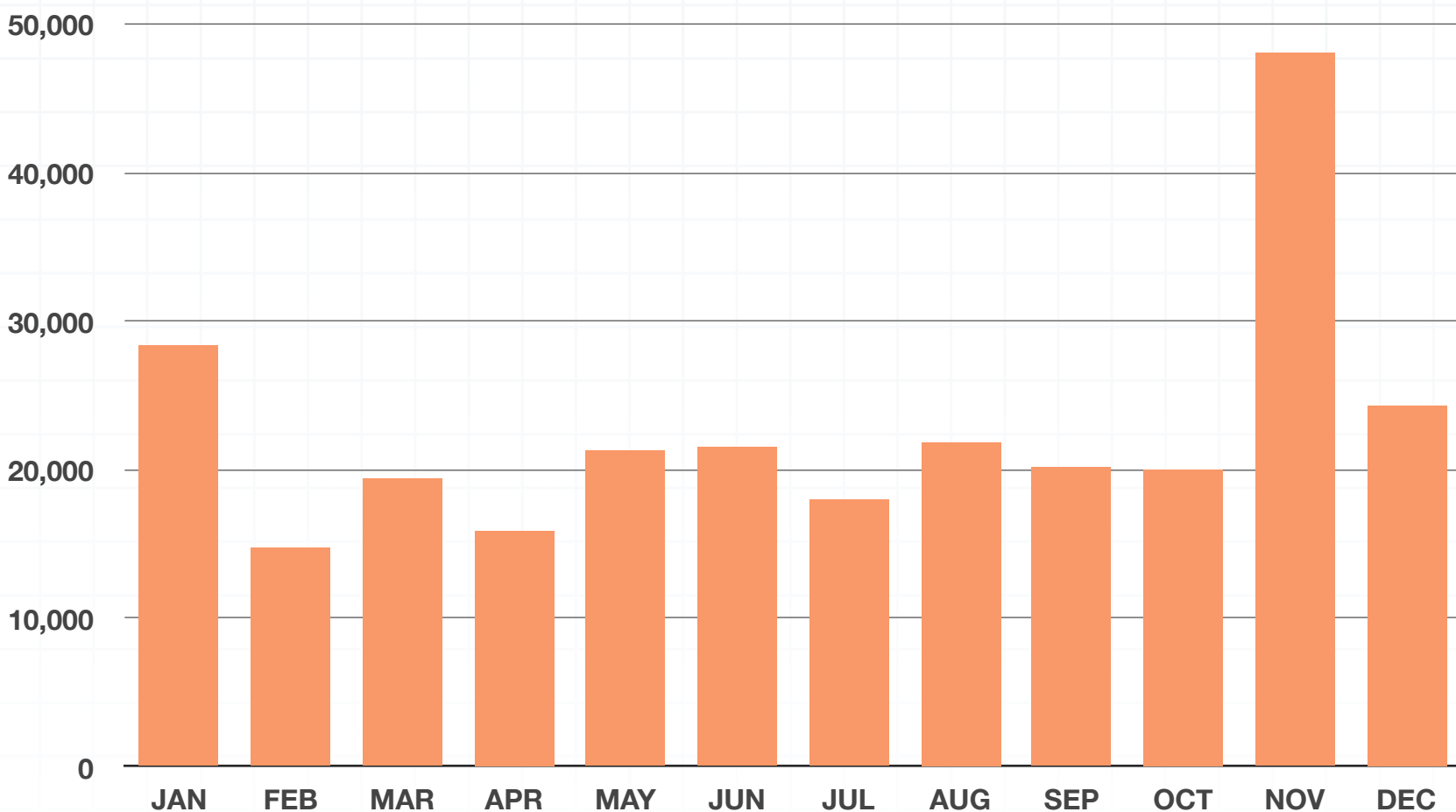
Month	Revenue
JAN	$28,361
FEB	$14,744
MAR	$19,407
APR	$15,891
MAY	$21,277
JUN	$21,530
JUL	$17,990
AUG	$21,838
SEP	$20,174
OCT	$20,025
NOV	$48,055
DEC	$24,318

In this one-way table, we have the months down the left side. Those are the individuals (also called the independent variable). The revenue is the variable (also called the dependent variable). In this chart, the revenue the artist earned from selling his paintings only depends on one thing: the month. We can only ask questions like "How much has the artist earned in February?" or "How much has the artist earned in September?"

Let's represent this as a bar graph.

This certainly gives us information about which months tend to be best for the artist. November and January seem to be best. But since the data is summarized, adding the revenue for January for all four years into one figure, and for February, for March, for April, etc., this one-way data doesn't give us any insight into how the artist's business has grown over time.

So to get more visibility, we could instead break apart the revenue information and show each year separately in **two-way table**:

Month	2014	2015	2016	2017
JAN	$254	$1,396	$10,696	$16,015
FEB	$377	$1,084	$4,745	$8,538
MAR	$291	$1,679	$7,549	$9,888
APR	$1,146	$1,668	$2,434	$10,643
MAY	$820	$2,708	$2,326	$15,423
JUN	$1,138	$3,014	$2,982	$14,396
JUL	$694	$2,586	$3,270	$11,440
AUG	$486	$2,080	$5,451	$13,821
SEP	$538	$3,109	$6,614	$9,913
OCT	$448	$3,695	$5,153	$10,729
NOV	$2,387	$6,495	$13,128	$26,045
DEC	$401	$2,030	$7,441	$14,446

Notice how the data is the two-way table is now dependent on two independent things, not just one. In the one-way table, if someone asked you how much revenue the artist earned, you'd reply "For which month?" But in the two-way table, if someone asked you how much revenue the artist earned, you'd reply "For which year and for which month?" In the two-way table, each value depends on two things: the month and the year. Whereas in the one-way table, each value depended on only on thing: the month.

We could graph the data in the two-way table in lots of different ways, but let's do it as a line graph that starts in January, 2014 and ends in December, 2017.
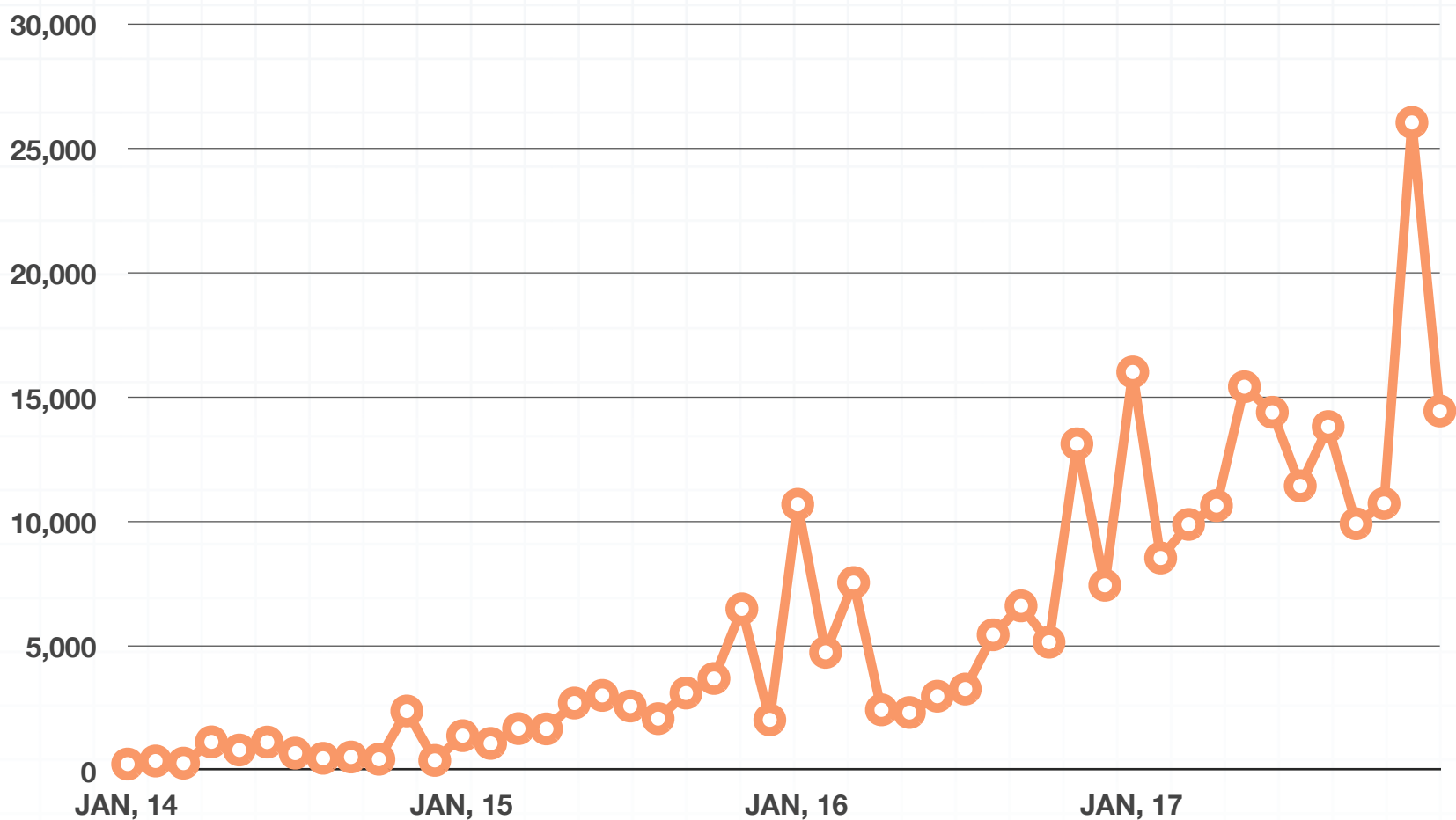
By giving the artist's revenue data in a two-way table, we get so much more insight into his business.

Showing totals

In two-way tables, we can also summarize the data by row and by column, plus give a total for the entire table in the lower right-hand corner.

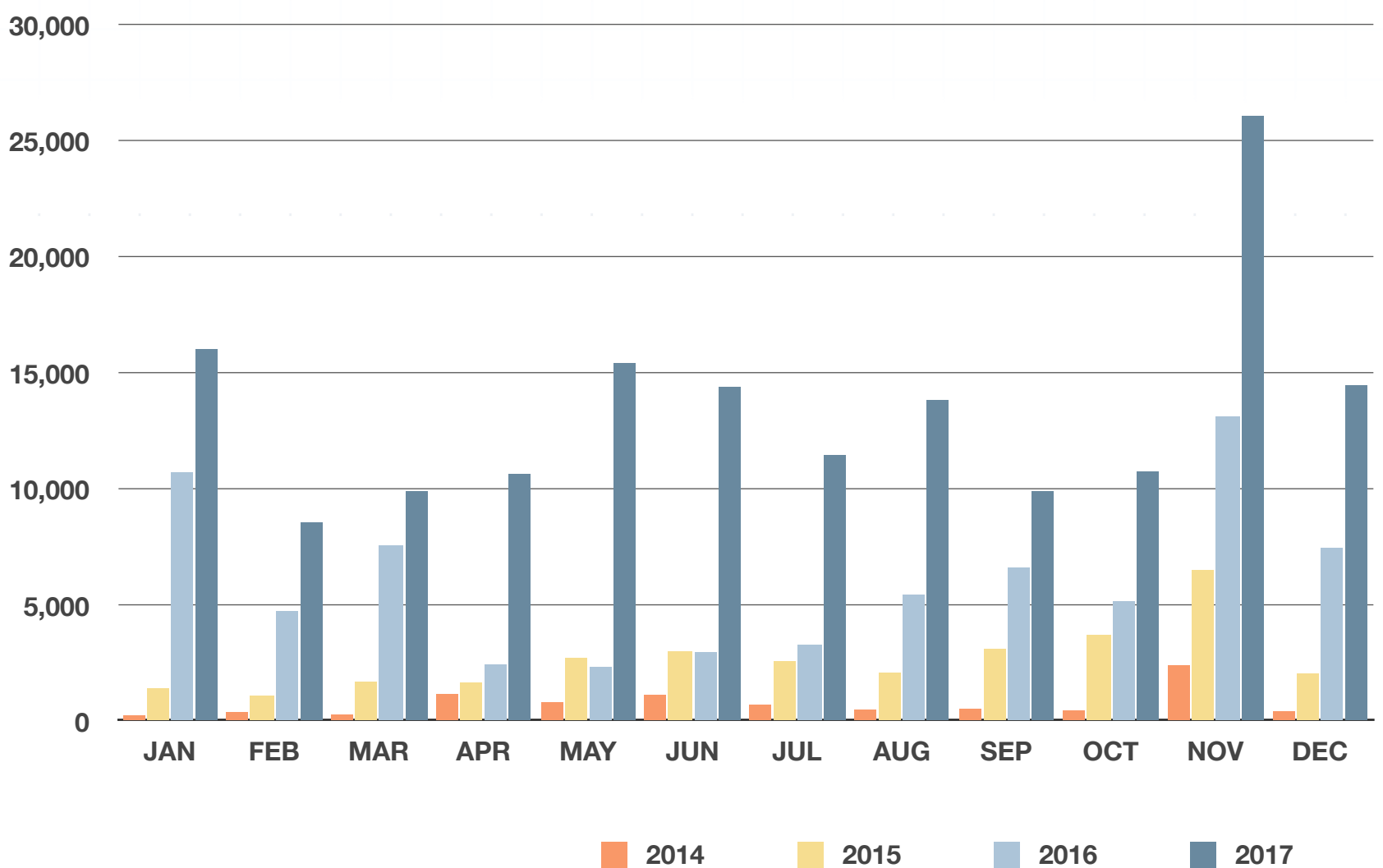| Month	2014	2015	2016	2017	Total
JAN	$254	$1,396	$10,696	$16,015	**$28,361**
FEB	$377	$1,084	$4,745	$8,538	**$14,744**
MAR	$291	$1,679	$7,549	$9,888	**$19,407**
APR	$1,146	$1,668	$2,434	$10,643	**$15,891**
MAY	$820	$2,708	$2,326	$15,423	**$21,277**
JUN	$1,138	$3,014	$2,982	$14,396	**$21,530**
JUL	$694	$2,586	$3,270	$11,440	**$17,990**
AUG	$486	$2,080	$5,451	$13,821	**$21,838**
SEP	$538	$3,109	$6,614	$9,913	**$20,174**
OCT	$448	$3,695	$5,153	$10,729	**$20,025**
NOV	$2,387	$6,495	$13,128	$26,045	**$48,055**
DEC	$401	$2,030	$7,441	$14,446	**$24,318**
Total	**$8,980**	**$31,544**	**$71,789**	**$161,297**	**$273,610**

From the two-way table, we can now see the revenue totals for each month down the right side of the table (these match the totals in the one-way table), but we can also see the revenue totals for each year along the bottom. We can still see which month is best for the artist (November), but we can also see which year was best for the artist (2017).

We did a line graph to represent all of this data over time, but we could have done a bar graph, instead.
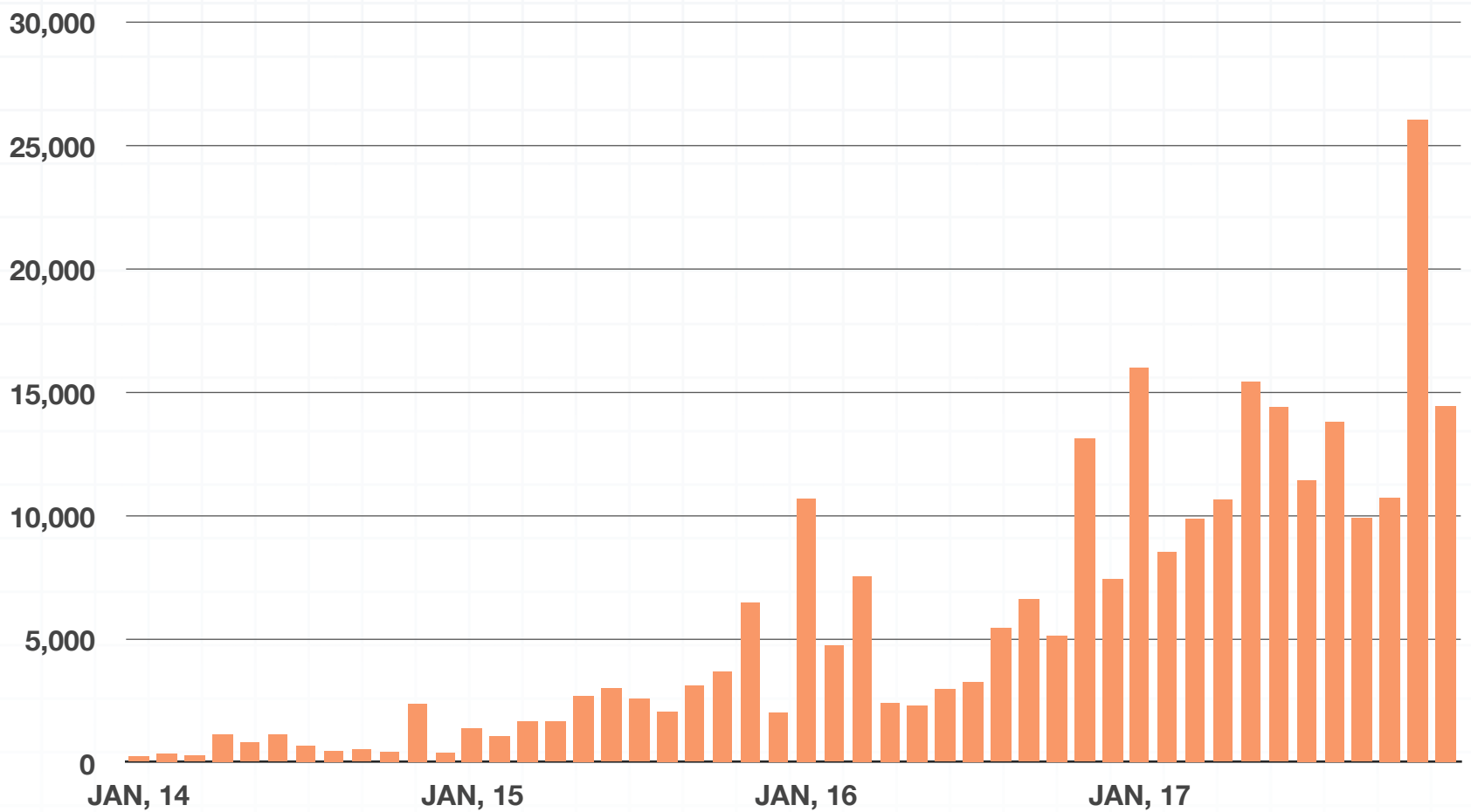
We could also create a **comparison bar graph**,

or a **comparison line graph**.