9.2.5 Generate the Summary

In addition to the plot we just made, we want to make sure to provide W. Avy with some solid statistical analysis—such as the mean, standard deviation, minimum, and maximum. He needs hard results if he's going to invest his money.

We're getting close to being able to deliver some of our findings to W. Avy.

REWIND

Here's a refresher on some key concepts in statistics:

- **Mean:** the average, which you can find by adding up all the numbers in a dataset and dividing by the number of numbers.
- **Variance:** how far a set of numbers is from the average.
- **Standard deviation:** a measure of how spread out the numbers in a dataset are; the square root of the variance.

- **Minimum:** the smallest number in a dataset.
- **Maximum:** the largest number in a dataset.
- **Percentiles:** where the number is in relation to the rest of the set of data.
- **Count:** the total number of numbers or items in a dataset.

Fortunately, Pandas helps us with these calculations. We'll use the `describe()` function to calculate the mean, minimum, maximum, standard deviation, and percentiles. Add the following to your code:

```
df.describe()
```

Now run everything you've got so far. The summary should look something like this:

	precipitation
count	2021.000000
mean	0.177279
std	0.461190
min	0.000000
25%	0.000000
50%	0.020000
75%	0.130000
max	6.700000

This data gives us a summary of different statistics for the amount of

precipitation in a year. The count is the number of times precipitation was observed. The other statistics are the precipitation amounts for each station for each day.

ADD/COMMIT/PUSH

You've accomplished a lot so far! Now is a good time to update or add your code to your GitHub repository.

Now that we've completed our precipitation analysis, we can share it with W. Avy and move on to the station analysis.