

## Mean

### Introduction

1 min

Finding the center of a dataset is one of the most common ways to summarize statistical findings. Often, people communicate the center of data using words like, *on average*, *usually*, or *often*.

In this lesson, you will learn how to calculate the *mean* of a dataset, a common measure of a dataset's center. We will use the mean to help us answer the question,

*When are adults their most creative and productive?*

You could define “creative” and “productive” in a lot of ways, making this question impossible to fully answer by the end of this lesson. However, you will form an informed opinion on the question using data of the one hundred greatest novels of all time.

We collected the dataset from a survey administered by the French literary magazine, Le Monde. From the dataset, you will calculate the average age of the authors when their books were published.

### Instructions

The histogram to the right displays the ages of 100 authors from the Le Monde survey. Where do you think the data is centered?

#### script.py

```
# Import packages
```

```
import codecademylib
```

```
import numpy as np
```

```
import pandas as pd
```

```
# Import matplotlib pyplot
```

```
from matplotlib import pyplot as plt
```

```
# Read in transactions data
```

```
greatest_books = pd.read_csv("top-hundred-books.csv")
```

```
# Save transaction times to a separate numpy array
```

```
author_ages = greatest_books['Ages']
```

```
# Use plt.hist() below
```

```
plt.hist(author_ages, range=(10, 80), bins=14, edgecolor='black')
```

```
plt.title("Age of Top 100 Authors at Publication")
```

```
plt.xlabel("Age")  
plt.ylabel("Count")  
  
plt.show()
```

