**Executive Summary**

**Overview**

The box-whisker plot illustrates the distribution of votes by year, highlighting the outliers. This visual representation is based on a dataset filtered to exclude rows with missing or non-numeric vote values, ensuring accuracy in the depiction of vote distribution.

**Data Preparation**

- The dataset is cleaned by removing rows with missing or non-numeric vote entries.

- The 'VOTES' column is converted to a numeric type to facilitate accurate plotting.

**Plot Characteristics**

- \*\**Palette*\*\*: The color palette "coolwarm" is used to differentiate between years.

- \*\*Outliers\*\*: Outliers are highlighted in red, making them easily distinguishable from the main distribution.

**Key Observations**

- \*\*Range and Variability\*\*: The range of votes varies significantly across different years, indicating fluctuations in voter turnout or recording over the years.

- \*\*Outliers\*\*: Specific years, such as 1960, 1991, and 2006, have numerous outliers, indicating exceptional cases of high voter turnout or reporting errors.

- \*\*Median Votes\*\*: The median number of votes, represented by the line inside each box, shows variation across years. For example, the year 2001 shows a higher median compared to earlier years like 1980.

**Numerical Insights**

- \*\*Years with Highest Variability\*\*: Years such as \*\*1991\*\* and \*\*2006\*\* exhibit high variability in votes, with the interquartile range (IQR) spanning a broad spectrum.

- \*\*Years with Most Outliers\*\*: Years like \*\*1960\*\* and \*\*2001\*\* have numerous outliers, indicating extreme values outside the typical vote distribution.

- \*\*Median Vote Counts\*\*: The median vote count has increased over time, with notable increases in years like \*\*2001\*\* and \*\*2021\*\*, where the median exceeds \*\*40,000 votes\*\*.

**Supporting Data**

- \*\*Years Included\*\*: The data spans the following years: \*\*1957\*\*, \*\*1960\*\*, \*\*1965\*\*, \*\*1967\*\*, \*\*1970\*\*, \*\*1977\*\*, \*\*1980\*\*, \*\*1982\*\*, \*\*1987\*\*, \*\*1991\*\*, \*\*1996\*\*, \*\*2001\*\*, \*\*2006\*\*, \*\*2011\*\*, \*\*2016\*\*, \*\*2021\*\*.

- \*\*Median Votes for Notable Years\*\*:

- \*\*1960\*\*: Approximately \*\*20,000 votes\*\*

- \*\*2001\*\*: Approximately \*\*40,000 votes\*\*

- \*\*2021\*\*: Approximately \*\*45,000 votes\*\*

- \*\*Years with Significant Outliers\*\*:

- \*\*1960\*\*: Several outliers above \*\*40,000 votes\*\*

- \*\*1991\*\*: Numerous outliers spread up to \*\*80,000 votes\*\*

- \*\*2006\*\*: Outliers reaching up to \*\*90,000 votes\*\*

**Conclusion**

The box-whisker plot effectively demonstrates the distribution of votes by year, with a clear highlight on outliers. This visualization can help in identifying trends, anomalies, and changes in voting patterns over the years. The presence of numerous outliers in specific years may warrant further investigation to understand the underlying causes, whether they are due to genuine increases in voter turnout or discrepancies in data collection.

**Recommendations**

- \*\*Further Analysis\*\*: Investigate the years with high outliers to understand the causes behind the extreme values.

- \*\*Data Quality Check\*\*: Ensure the data collection methods are consistent across years to minimize discrepancies.

- \*\*Trend Analysis\*\*: Use the data to analyze trends in voter turnout and correlate with historical events to gain deeper insights.

This summary encapsulates the key findings from the box-whisker plot, providing a clear and concise overview of the data distribution and notable patterns.