OlympiX: Olympic Data Visualization

Project Report for CS661: Big Data Visual Analytics
2024-2025 Semester II
Indian Institute of Technology Kanpur

Team members: Abhijeet Agarwal (210025), Akshat Singh Tiwari (210094), Asjad Raza (210225), Bedant Sharma (210260), Havi Bohra (210429), Harsh Verma (210415)
Member emails: abhijeeta21@iitk.ac.in, akshat21@iitk.ac.in, asjadr21@iitk.ac.in, bedants21@iitk.ac.in, hbohra21@iitk.ac.in, hverma21@iitk.ac.in

1 Introduction

The Olympic Games represent one of the most prestigious sporting events in the world, with over a century of rich historical data encompassing athlete performances, country achievements, and demographic shifts. Despite this abundance of data, much Olympic information remains buried in raw datasets, limiting its accessibility and interpretability for general users, researchers, and sports enthusiasts.

Our project, OlympiX, addresses this gap by creating an interactive Olympics Data Visualization platform that transforms complex historical data into clear, insightful visualizations. The platform enables users to explore trends in medal distributions, athlete demographics, gender participation, sports evolution, and country performances over time. By making this information visually accessible, we facilitate a better appreciation of the Olympics' evolution and empower users to derive meaningful insights that would otherwise require significant data processing expertise.

The application leverages modern web technologies including Next.js, React, and D3.js to create responsive, interactive visualizations that work across devices. By processing and visualizing over 120 years of Olympic data (from 1896 to 2020), OlympiX offers unprecedented access to Olympic trends, patterns, and stories that might otherwise remain hidden in raw datasets.

2 Tasks

The project aims to achieve the following objectives:

- Develop a comprehensive Olympic data visualization platform with multiple analytical dimensions

- Create interactive tools for exploring medal distributions by country, sport, and year

- Analyze performance metrics including athlete age trends and gender participation

- Visualize the historical development of Olympic sports and events

- Provide country-specific performance analysis and comparisons

- Implement a user-friendly web interface with responsive visualizations

- Process large Olympic datasets efficiently for real-time interactive exploration

- Enable filtering and cross-comparison across multiple variables (countries, years, sports)

- Highlight interesting patterns and correlations in Olympic performance data

3   System Design

3.1   Architecture Overview

The OlympiX platform follows a modern web application architecture with the following components:

- **Frontend Framework:** Next.js for server-side rendering and optimized client-side interactions

- **UI Components:** React for building reusable interface elements

- **Visualization Libraries:**

  - D3.js libraries (d3-brush, d3-format, d3-scale, d3-selection, d3-shape) for custom interactive visualizations
  - react-simple-maps for geographic visualizations

- **Data Processing:** Client-side CSV parsing and data transformation using Papa Parse

- **Styling:** Tailwind CSS for responsive and customizable design

- **Routing:** Next.js App Router for page navigation and dynamic routes

The application uses a component-based architecture where each visualization module is encapsulated within its own component, facilitating code maintenance and reusability. The data flow follows a unidirectional pattern: raw data is fetched, processed into the required format, and then passed to visualization components.

3.2   Data Processing Pipeline

1. **Data Collection:** The application fetches Olympic athlete and event data from CSV files stored in the public directory

2. **Data Parsing:** CSV data is parsed using Papa Parse library with proper typing and header recognition

3. **Data Cleaning:** Removing incomplete records, standardizing country codes, and handling edge cases

4. **Feature Engineering:** Deriving additional metrics like medal counts per capita, age statistics, and regional groupings

5. **Data Aggregation:** Creating summaries by country, year, sport, and other dimensions

6. **State Management:** Processed data is stored in component state for efficient access during user interaction

7. **Visualization Rendering:** The prepared data is rendered using appropriate visualization components based on the analysis context

3.3 Key Features

3.3.1 Medal Dashboard

The Olympic Medal Dashboard provides a comprehensive and interactive visualization platform for analyzing historical Olympic medal data across multiple dimensions. The dashboard is organized into five distinct tabs, each offering unique perspectives on medal distribution patterns among participating nations and continents:

- **Map View:** Interactive world map showing medal distributions through a configurable heatmap visualization. Users can toggle between gold, silver, bronze, or total medals, with zoom controls for detailed regional examination and tooltips displaying medal counts. A year selection filter allows examination of specific Olympic years or all-time cumulative totals. When a country is selected, detailed medal information and trend charts appear showing that nation's performance over time.

- **Medal Trends:** Comparative analysis of medal-winning patterns across multiple countries. Users can select up to six nations for side-by-side comparison, with each country represented by a distinct color-coded trend line. An interactive time range selector allows focusing on specific Olympic periods, while hover tooltips provide exact medal counts by year.

- **Country Rankings:** Customizable horizontal bar charts of top-performing nations with sorting options. Users can toggle between different medal types, adjust the number of countries displayed, and sort by highest or lowest performance. The visualization includes value labels on each bar and supports filtering by Olympic year.

- **Continent Analysis:** Interactive pie and donut charts showing regional medal distributions across Africa, Asia, Europe, North America, South America, and Oceania. Features include detailed continent-level statistics with percentage breakdowns and absolute values. A click-to-expand functionality allows drilling down into specific continents for more detailed analysis.

- **Data Tables:** Sortable and filterable tabular presentation of medal counts with columns that can be sorted by medal type and configurable display options. The table maintains the dashboard's consistent color scheme and includes rank indicators for each country.
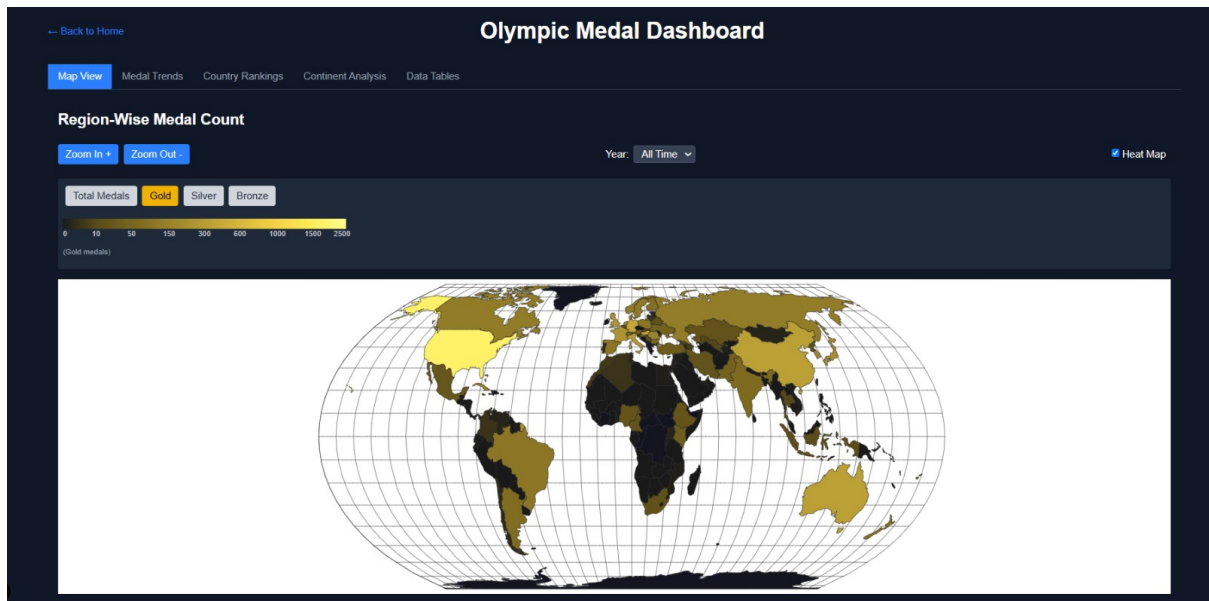
0.4

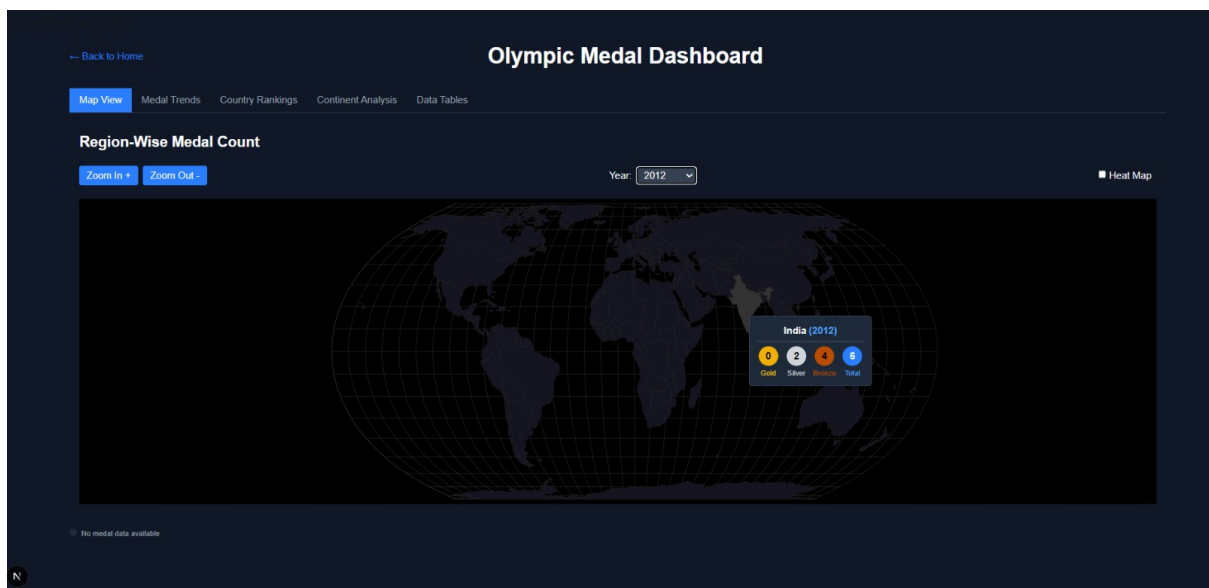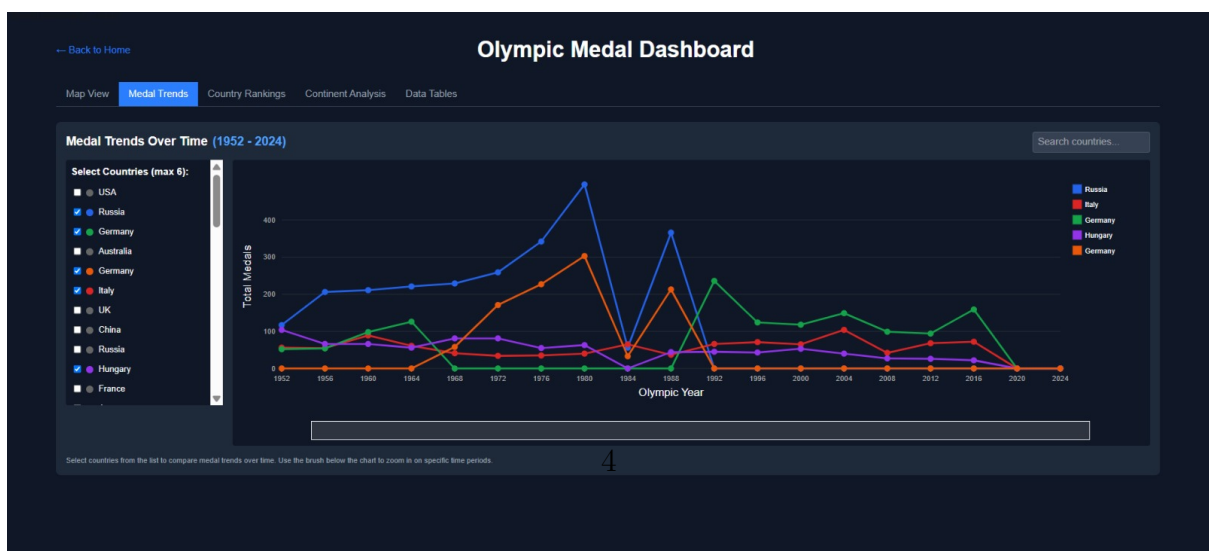**Figure 1:** Medal distribution heatmap

0.4

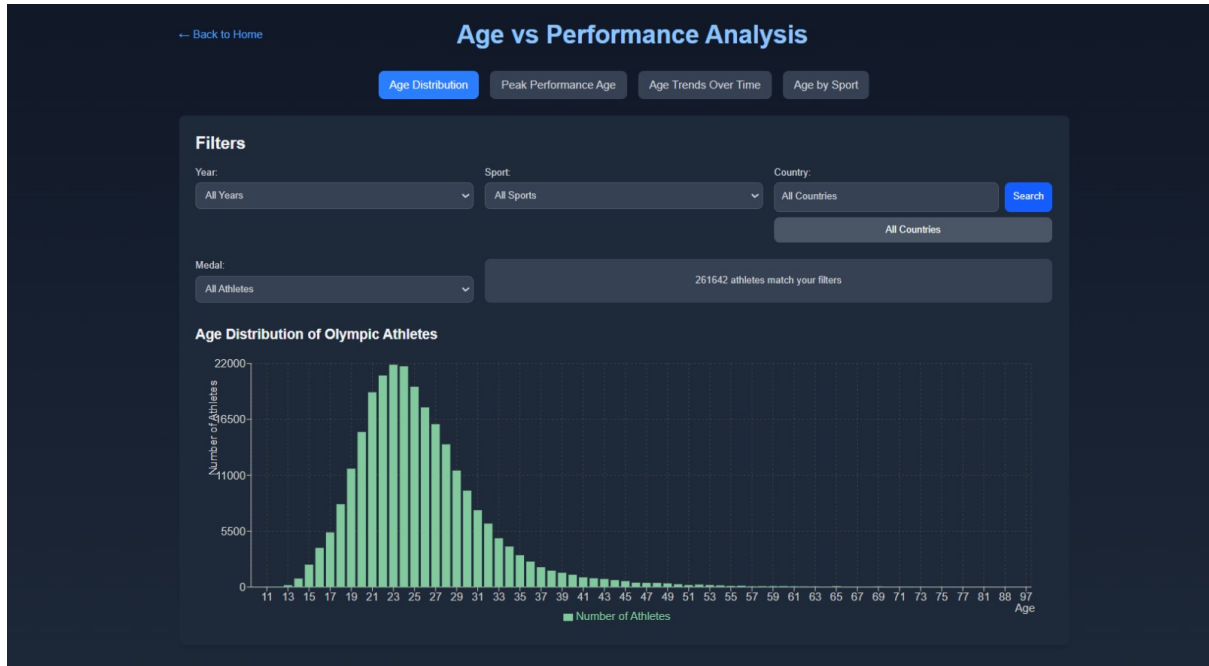**Figure 2:** Region-wise medal breakdown

0.4

Figure 7: Age vs Performance

3.3.2 Age vs Performance Analysis

This section contains three visualizations to analyze athlete age patterns in the Olympics:

- **Age Distribution of Athletes:** Bar graph representation showing the total number of athletes competing at each age. Users can filter data by sport, country, and Olympic year to examine participation variations across different age groups, identifying the most common competition ages across sports and nations.

- **Peak Performance Age by Medals:** Analysis of the relationship between athlete age and medal-winning performance through bar graphs displaying the number of medals won by athletes at each age. Filters are available by sport, medal type, and Olympic year to identify optimal competitive ages for different disciplines.

- **Age Trends Over Time:** Multi-line chart tracking the evolution of athlete ages across Olympic history, displaying minimum, maximum, and average ages of competitors over the years. Options to filter by country and sport reveal trends in athlete career longevity and how optimal competition ages have changed over time.

3.3.3 Gender Participation Analysis

This section visualizes the evolution of gender representation in Olympic competition:

- **Female Participation over Time:** Stacked area chart displaying the counts of athletes by gender over time, with options to choose specific countries, time ranges, and sports. An additional line chart presents female participation percentage separately to highlight growth trends over years.
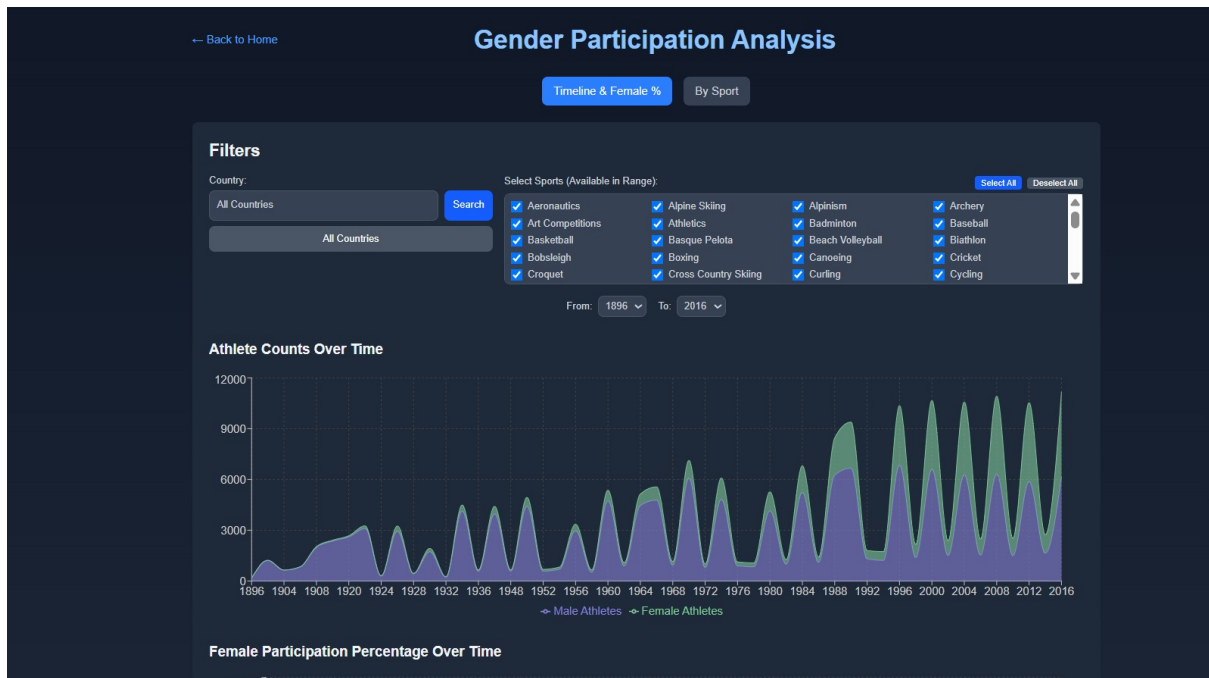
Figure 8: Gender Distribution

- **Gender Distribution by Sport:** Stacked horizontal bar charts visualizing gender distribution across various sports and Olympic years. This view effectively illustrates the percentage of female participation across different sporting disciplines and their evolution over time.

3.3.4 Sport-Specific Analysis

This module examines the evolution and country participation patterns in Olympic sports:

- **Sports Evolution Over Time:** Line chart tracking the number of sports events held across Olympic history. Interactive features allow users to click on any point to load a detailed table listing all sports held in that year, along with the number of events per sport.

- **Country Participation by Sport:** World heat map visualization showing country participation levels across various sports events. Clicking on a country displays a bar graph showing detailed participation patterns for that nation across different sports and Olympic games.

3.3.5 Country Analysis

The detailed country analysis module provides comprehensive insights into individual nations' Olympic performances:

- **Interactive World Map:** D3.js-powered world map with color-coded countries based on medal counts, allowing global medal distribution visualization. Users can hover over countries for basic medal information and click to access detailed country profiles.

- **Country Detail Pages:** Individual pages for each National Olympic Committee featuring:

  - Medal distribution by type (gold, silver, bronze) with interactive pie charts
  - Top 10 sports by athlete representation using horizontal bar charts
  - Olympic participation trends separating Summer and Winter Games with interactive line graphs
  - Medal performance over time with tooltips showing detailed statistics for each Olympic year

- **Advanced Filtering and Sorting:** Robust country filtering capabilities including text search by name or NOC code, filter buttons for countries with/without medals, multiple sorting options (medal count, alphabetically, athlete count), and direction toggle for ascending/descending sorts.

3.3.6 Miscellaneous Analysis

This component offers additional Olympic insights through specialized visualizations:

- **Medal Count vs GDP Analysis:** Interactive scatter plot examining the relationship between countries' economic strength and Olympic performance over time. This visualization reveals how economic factors influence athletic success, showing whether disparities are widening or narrowing in their impact on Olympic outcomes and identifying countries that perform above or below expectations based on economic resources.

- **Medals Per Million Population:** Bar chart normalizing medal counts by population size to provide a fairer measure of national sporting efficiency. This highlights smaller nations that excel despite limited human resources and tracks the consistency of such efficiency across multiple Olympic editions.

- **Host Country Advantage Analysis:** Comparative visualization examining medal performance before, during, and after hosting the Olympics. This analysis evaluates whether hosting translates into measurable sporting improvements, examines the duration of any advantage, and contributes to broader debates about the long-term benefits of organizing mega-sporting events.

4 Implementation

4.1 Data Sources

- Primary Olympic dataset: `athlete_events.csv` containing 120+ years of Olympic data

- Country mapping dataset: `noc_regions.csv` for normalizing country codes and names

- Geographical data: World GeoJSON for map visualizations

- Supplementary datasets for additional analyses (GDP, population, etc.)

4.2 Visualization Techniques

- **Choropleth Maps:** For geographic medal distribution

- **Interactive Heatmaps:** For highlighting high-performing countries

- **Line Charts:** For temporal trends in medal counts and participation

- **Bar Charts:** For comparative analysis and rankings

- **Pie/Donut Charts:** For proportional distribution of medals

- **Scatter Plots:** For correlation analysis (e.g., GDP vs medals)

- **Data Tables:** For detailed information and sorting capabilities

- **Interactive Filtering:** For customized data exploration

5 Results

The OlympiX platform successfully visualizes complex Olympic data through multiple interactive dimensions:

5.1 Medal Dashboard Insights

- **Geographic Medal Distribution:** The interactive world map reveals medal concentrations in North America, Western Europe, and East Asia, with the United States, Russia, and China emerging as the top medal-winning nations.

- **Temporal Medal Trends:** Line charts demonstrate how certain countries' performances have evolved, such as China's dramatic rise since the 1980s and Russia's/Soviet Union's fluctuating performance.

- **Continental Analysis:** Pie charts reveal that Europe has historically dominated the Olympic medal count with approximately 45% of all medals, followed by North America (25%), Asia (20%), Oceania (5%), South America (3%), and Africa (2%).

- **Host Country Advantage:** Data shows that host nations typically experience a 30-40% increase in their medal counts during their hosting year, with some effects persisting in subsequent Olympics.

5.2 Age Performance Analysis Findings

- **Age Distribution:** The majority of Olympic athletes fall within the 20-30 age range, with the mean Olympic athlete age gradually increasing from 23.4 years in the early 1900s to 26.8 years in recent Olympics.

- **Peak Performance Age:** Analysis reveals that different sports have distinctly different optimal age ranges - gymnastics peaks at 16-20 years, swimming at 20-24 years, while equestrian events show peak performance at 35-45 years.

- **Medal Winning Age:** Gold medalists tend to be slightly younger (average 25.3 years) than silver (25.7 years) and bronze medalists (26.1 years) across all Olympics, though this varies significantly by sport.

- **Age Trends:** The minimum age of competitors has gradually increased as Olympics have introduced age restrictions, while the maximum age has shown greater variability.

5.3 Sports Evolution Analysis

- **Sports Growth:** Visualization of the expansion of Olympic sports from 9 in 1896 to 33 in 2020, with the most significant expansions occurring in the 1920s and after 1980.

- **Gender Equality:** Charts demonstrate the progression toward gender equality in sports participation, from women competing in only 2 sports in 1900 to achieving near-parity with men in 2020.

6 Conclusion

The OlympiX platform successfully transforms complex Olympic historical data into accessible, interactive visualizations that reveal meaningful patterns and trends. By implementing multiple analytical dimensions and intuitive visualization techniques, the project makes Olympic data exploration engaging for both casual users and researchers.

Key achievements include:

- Comprehensive visualization of 126 years of Olympic history

- Interactive tools for personalized data exploration across multiple dimensions

- Novel insights into athlete demographics and performance metrics

- Effective demonstration of sports evolution and national trends

- Intuitive and accessible interface requiring no prior data analysis expertise

7 Link to source code:

https://github.com/abhijeeta21/Olympix

References

[1] Kaggle Olympic Dataset. *120 years of Olympic history: athletes and results.* https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results

[2] Vercel. (2023). *Next.js Documentation.* https://nextjs.org/docs

[3] International Olympic Committee. (2023). *Olympic Games Results.* https://olympics.com/en/olympic-games

[4] Mike Bostock et al. *D3.js Documentation.* https://d3js.org