**NHL Data Project**

**Overview**

This project downloads, processes, and visualizes NHL play-by-play data for seasons 2016–2023. It produces clean datasets and an interactive tool to explore games and events.

**Step-by-Step Process**

**Step 1: Data Acquisition**

**Purpose:** Download raw NHL game data from the API.  
**What it does:**

* Connects to NHL API endpoints
* Downloads play-by-play data for seasons 2016–2023
* Saves raw JSON files for each game

**Files created:**

* data/raw/season\_XXXX/general/
* data/raw/season\_XXXX/playoff/

**Run:**

python download\_all\_data.py

**Step 2: Data Processing**

**Purpose:** Convert raw JSON files into structured data.  
**What it does:**

* Parses JSON game files
* Extracts shots and goals events
* Creates clean CSV files for each game
* Adds player names and team information

**Files created:**

* data/processed/season\_XXXX/general/
* data/processed/season\_XXXX/playoff/

**Run:**

python data\_processor.py

**Step 3: Tidy Data Creation**

**Purpose:** Combine all games into analysis-ready datasets.  
**What it does:**

* Combines all individual game files
* Creates unified plays.csv and players.csv
* Calculates additional features (distance, shot angles)
* Cleans missing values

**Files created:**

* data/tidy/plays.csv
* data/tidy/players.csv

**Run:**

python tidy\_data\_creator.py

**Step 4: Data Wrangling**

**Purpose:** Final data preparation and basic analysis.  
**What it does:**

* Creates summary statistics
* Generates basic visualizations
* Prepares data for interactive tool

**Files created:**

* data/tidy/all\_seasons\_combined.csv

**Run:**

python data\_wrangling.py

**Step 5: Interactive Visualization**

**Purpose:** Explore data with interactive widgets.  
**What it does:**

* Creates dropdown menus for seasons and games
* Shows individual events on NHL rink
* Displays shot maps for entire games
* Provides detailed event information

**Run in Jupyter Notebook:**

from interactive\_tool import launch\_interactive\_tool

launch\_interactive\_tool()

**Expected Output**

After completing all steps, the project will produce:

**Data Files:**

* data/raw/ — Raw JSON game files (~20,000+ files)
* data/processed/ — Clean individual game CSVs
* data/tidy/ — Combined analysis datasets

**Interactive Tool Features:**

* Season selection (2016–2023)
* Game type (Regular season / Playoffs)
* Game browser with event slider
* Real NHL rink visualization
* Shot and goal plotting with coordinates
* Detailed event information