--- Group B Final project – data part

Slide 1: -- Data description

Thanks Christina.

Thank you everyone for joining us. Let’s review the data.

This report uses Kaggle Dataset “120 Years of Olympic History: Athletes and Results”.

This dataset provides 136K observations on 15 variables. In preparatory data processing, the following variables were selected for analysis: Sex, Age, Team, Year, Season, Medal and Country.

After analyzed different aspects in the whole Olympic event data, only subset data from 2000 were chosen for better understanding the sport trends and player status in different countries.

There are total 5024 observations in the subset data.

Data were extracted, processed, analyzed with R. The subset data includes 5 summer Olympic events from 2000 to 2016, and 4 Winter Olympic events from 2002 to 2014. We analyzed players under 19 year old, both male and female, from all countries, but we focus on medal winning players NOT from top 10 countries.

Slide 2: all player info

-- Methods/ Approach: A description of the techniques or algorithms to solve the problem.

Data preprocessing includes extracting interested data, such as aggregate, sort, join, merge and append data frames. Line charts and bar chart were produced for better understanding and easy comparison. There are 2 different charts for comparison, for example player in summer Olympics vs winter Olympics, or the number of players vs the number of player winning medals.

The charts in this slide indicate the number of players under 19 years old from around the world is decreasing in summer Olympic. But there is no big change in the number of winter Olympics.

Slide 3: top 10 vs non-top 10

The charts in this slide show the total number of players under 19 years old from in non-top 10 medal winning countries. The players number is decreasing from 2000 summer Olympic.

Slide 4: male vs female

This slide show the number of male and female players under 19 years old from medal winning countries in non-top 10 medal winning countries.

It also indicates that there are a greater number of females under 19 participating than males in the same demographic. The number in summer Olympic is going down dramatically. Athletes of this age group could be under resourced and good candidates for sponsorships.

Slide 5: USA player male vs female

This slide shows the number of U.S. players under 19 are similar, female athletes far more than male athletes.

Next, Sujoy will highlight the results and talk about the future plan.