**Smoking Trend - Design Report**

Team 5

University of Washington

IMT 562 - Interactive Data Visualization

Professor/Dr. Phil Fawcett

03/16/2020

## Abstract

This report is used to record the exploration of the smoking-related datasets, answer questions based on the data, conduct usability tests, record any refinements of the visualization design based on usability test results, and summarize our findings and recommendations. Also, it contains questions used in usability tests, information about the participants, and analysis based on the usability test results.

## Introduction

This interactive visualization, smoking trend, is meant to provide users a means to explore and analyze how the trend of smoking cigarettes has changed over 33 years from 1980 to 2012. It takes multiple facets for a comprehensive analysis. Other than the most straightforward aspect, the total consumption of cigarettes, the visualization takes other approaches like smoking prevalence in females and males, per smoker per day, and the number of smokers. Within each facet, we look specifically at gender distributions, countries, ages, and trends. In this way, we can get a holistic view on the data and rule out outliers, if there is any, such as people who purchase cigarettes for sale instead of smoking themselves, in order to provide a more precise analysis tool. To make our visualization more accessible and user-friendly, we conduct 5 usability tests to study the flaws in our visualization and refine it based on the feedback. The participants in the usability tests would be asked to think out loud so that we would not miss any small details in the thinking process.

## Area of Focus

We gathered a group of datasets from the Institute of Health Metrics and Evaluation (IHME) about smoking from 1980 to 2012. The datasets are originally from a study conducted by an iSchool Professor, Michael Freeman, at the University of Washington. The datasets are on total consumption, the number of smokers, total consumption per smoker per day, and smoking prevalence of females, males, and a combination of them. Our focus is to show the patterns or trends of smoking in the aspects mentioned before globally so that readers will be able to see how the trend of smoking has changed and impacted the world. Furthermore, we focus to enable readers to analyze various aspects of smoking for one specific country. Even if the trend does not necessarily indicate which factor is influencing the country, it can provide readers some possible guesses. For example, we discover an obvious decrease in total cigarette consumption. This might suggest, but not necessarily prove, the impact of alternative choices, such as vaping and hemp. In sum, our focus is to find out the trends of smoking and some underlying factors that have caused those trends.

## Summary of Question

Our main question that would be answered by the data visualization is “how has the trend of smoking changed throughout 33 years from 1980 to 2012? ”. There are some following questions that our visualization answers to prompt a further and deeper look into our main question. For example, which country/gender/age has more smokers than another country? Or who is devoting to the increase in the trend of smoking?

## Motivation

Two reasons motivate our team to learn more and search for smoking-related dataset. One of the reasons is that we find out there are much less people smoking in the U.S compared to China. This is what we have realized by observing in life, but we want to use actual data to prove the real trend of smoking in these countries. The other reason is due to the increasing concerns for coronavirus, which has caused us to pay much more attention to health issues, especially pneumonia. Pneumonia is a disease caused by the coronavirus. When we were researching the main factors that cause pneumonia, we found smoking to be one of the largest factors. Therefore, we went on to search for datasets on smoking, and we found our datasets on IHME. We believe that this is an important topic to analyze because smoking has been negatively affecting so many people in the world. It does not only hurt the lungs, but also a lot of other organs. It’s important to show the world the increase in smoking so that the governments in every country will attach importance to these issues and maybe start legislation to control smoking, such as increasing taxes for cigarettes.

## Interactive Visualization

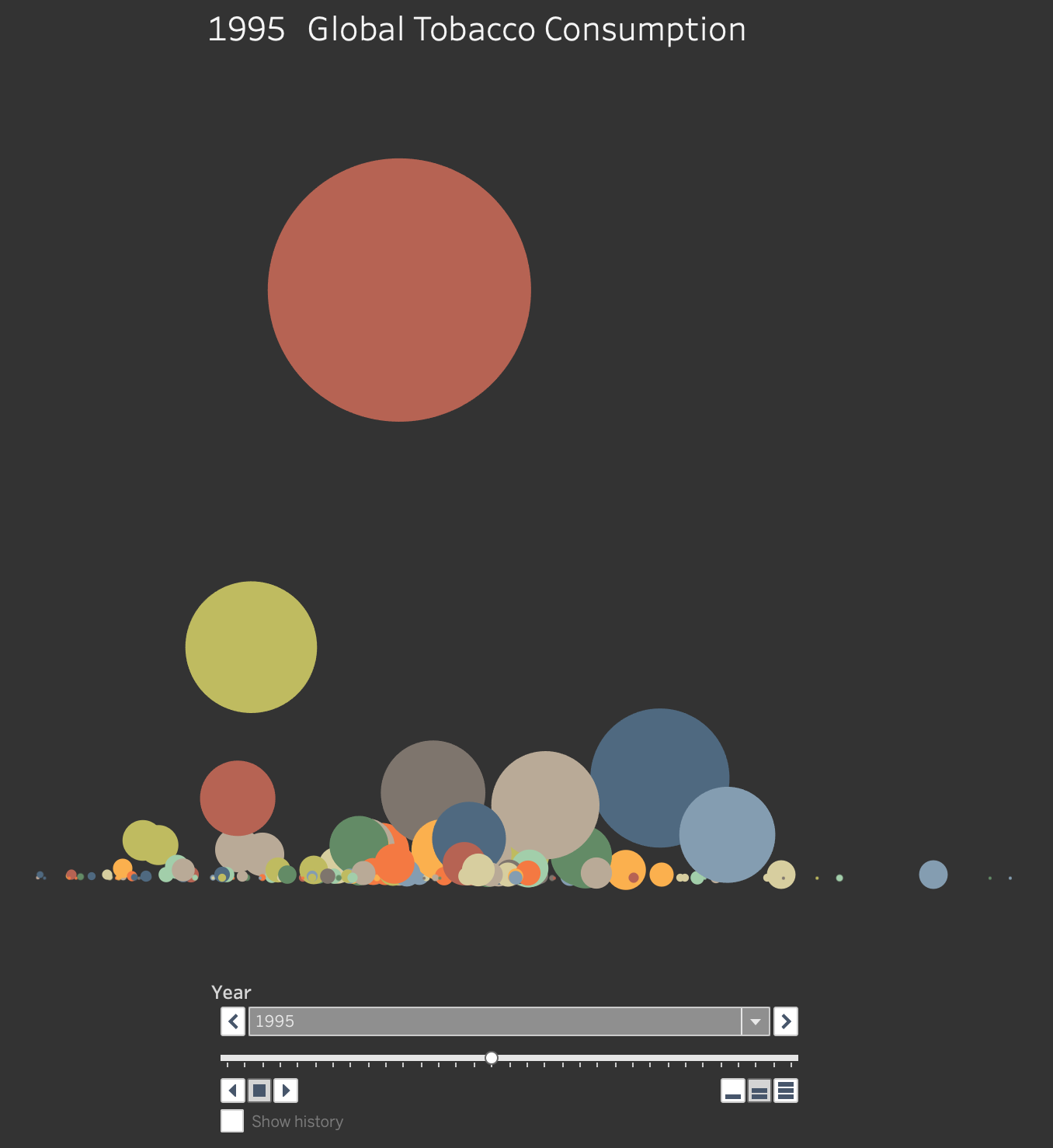
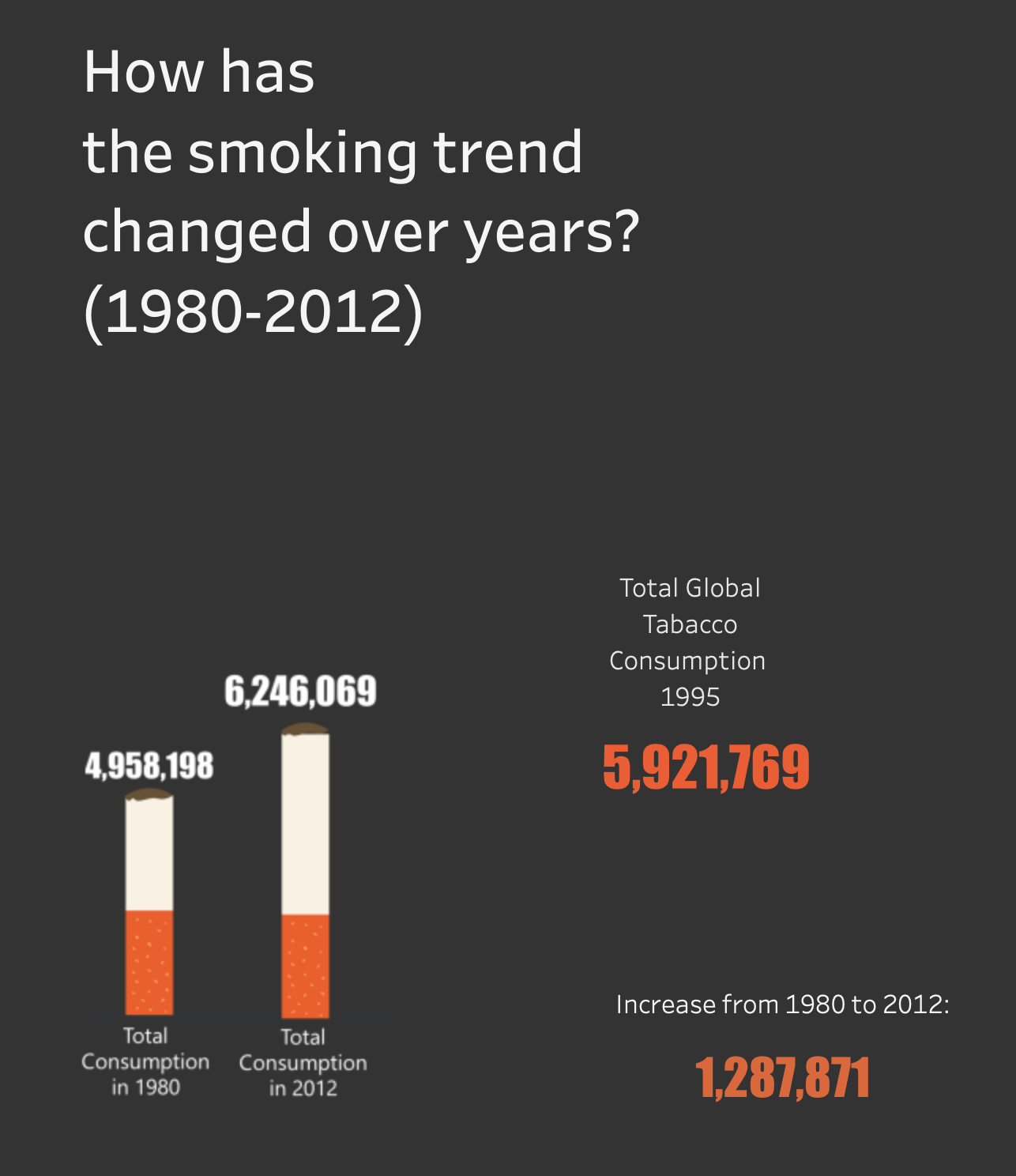


Figure 1. Global Total consumption

Figure 1. is a bubble graph on the front page. The y-axis is the total consumption per smoker per day. The x-axis is the number of smokers. The size of bubbles is the total consumption. We intentionally do not mark the x and y axes because we want to leave a hook here to attract and motivate readers to continue exploring what the data is about and what trends it is delivering. With the guiding question at the left, users will continue the exploration easier. The big number at the left about the total global consumption will change corresponding to the year filter.

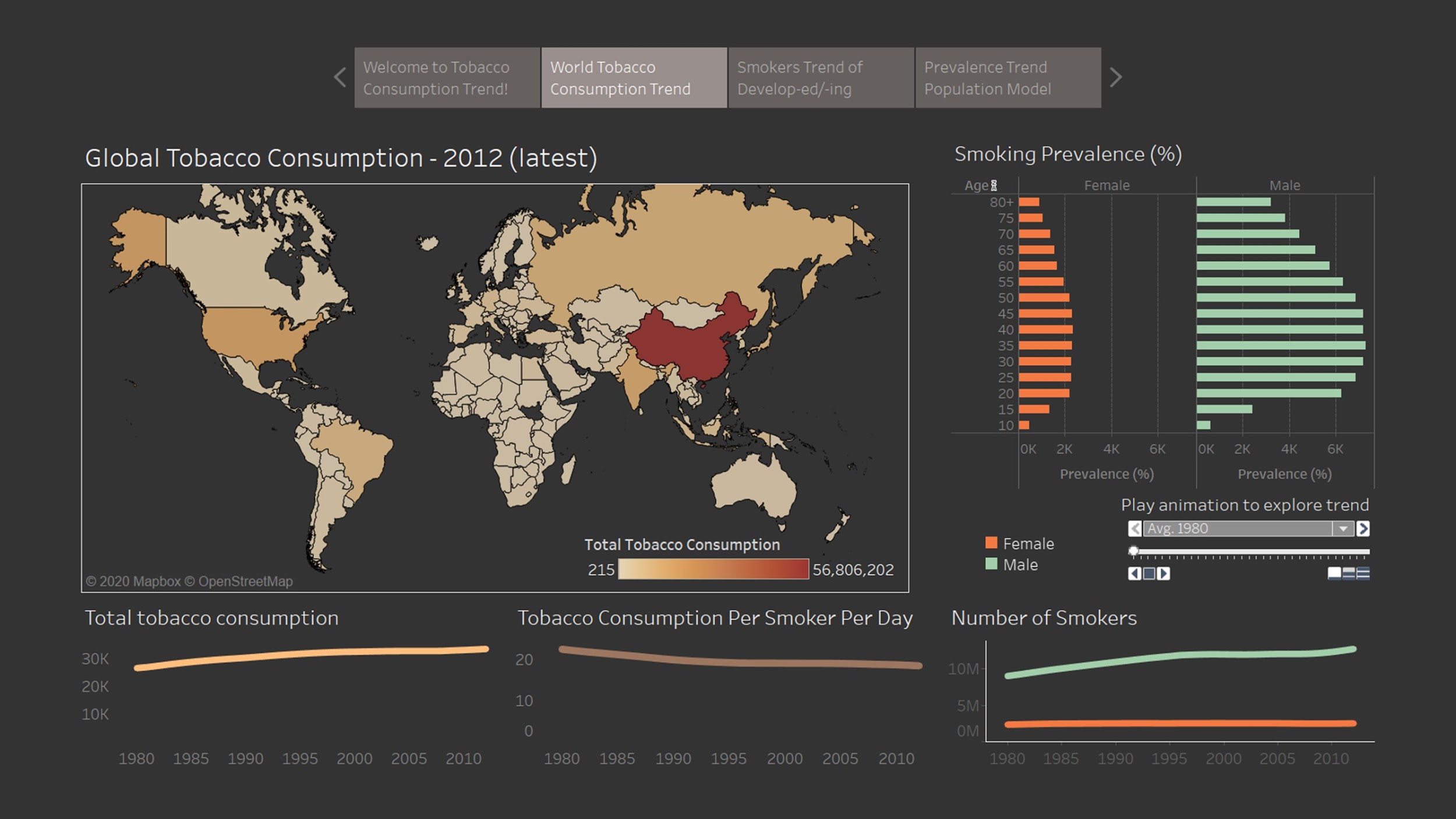


Figure 2. World cigarette consumption trend

Figure 2. is an overview of the global trend of smoking. In the initial state when there is no country selected, the graphs all present the global average value for each facet. The map shows the total consumption in the latest year, which is 2012. The prevalence bar charts show the percentage of smokers in the total population separately for females and males. The animation enables readers to see the changes in prevalence across the years. The three line graphs show the changes in the trend of smoking in total consumption, consumption per person per day, and the number of smokers across years. These graphs give an overall idea of how the trend of smoking changed across the years. The bar graph and the line graphs can be filtered into a country by clicking on a country on the map.

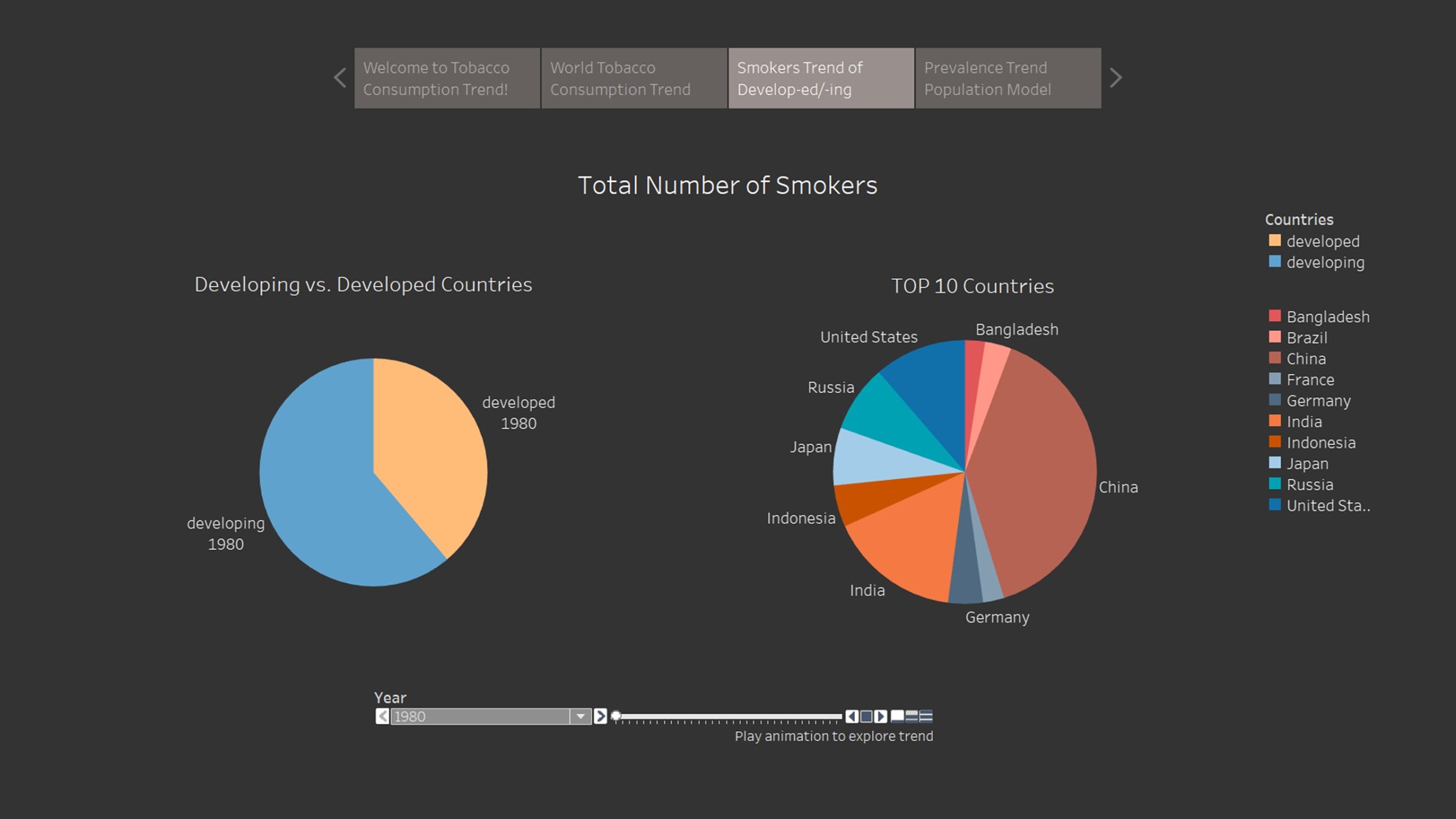


Figure 3. Smoker trends in developed/developing countries

Figure 3. shows the total number of smokers in pie charts. One is meant to show the difference between developed and developing countries across the years. The other is meant to show the ratio of smokers in each country across the years.

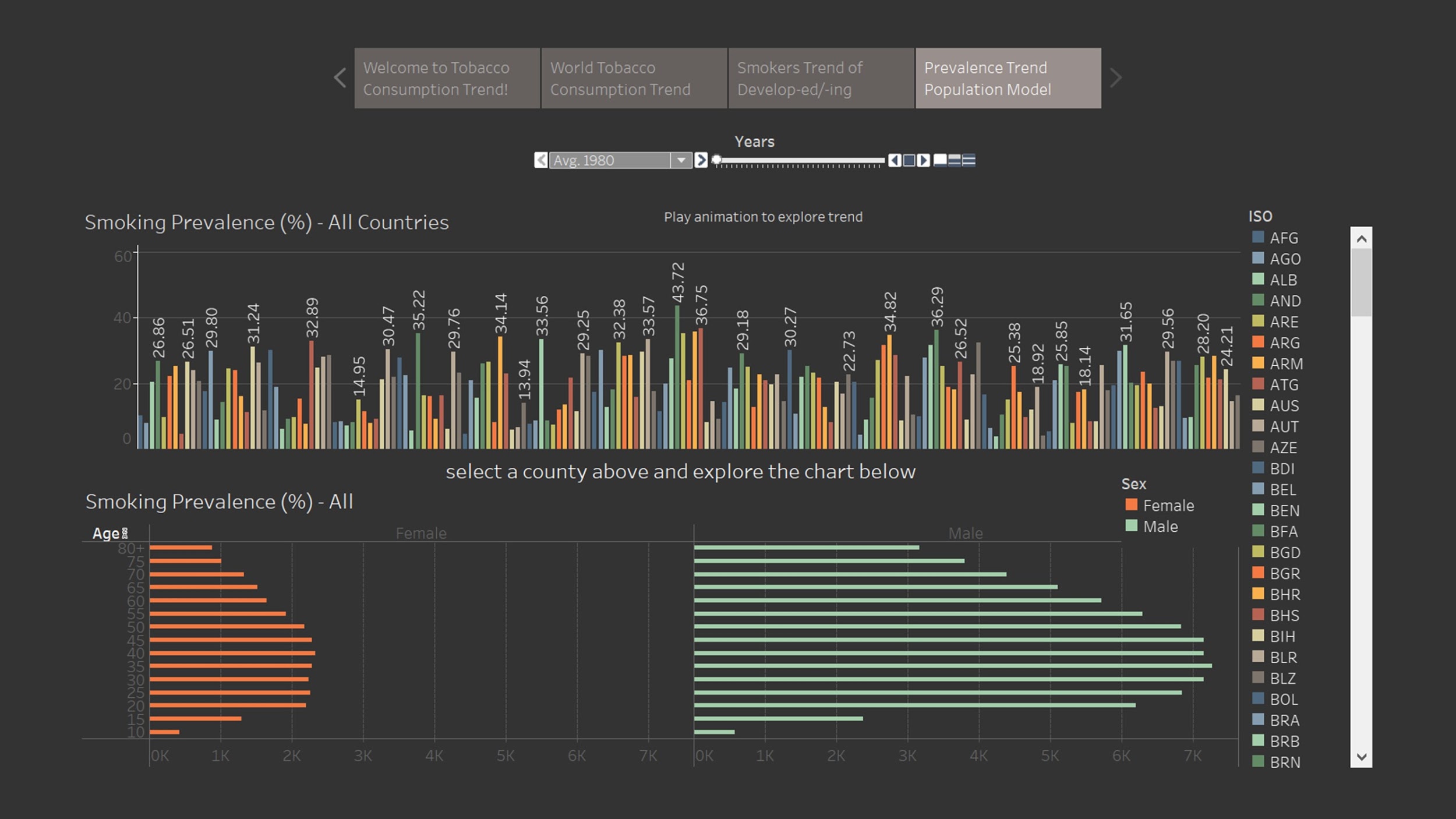


Figure 4. Prevalence trend models

Figure 4. is about exploring the changes in prevalence by country, age, and sex. The animation enables users to see the changes in the trend of smoking prevalence across the years. By clicking on a bar or a country on the legend, the graph at the bottom will be filtered to be about that country. This bottom graph is based on age and sex. If we click on anywhere on the bottom graph, the corresponding country will be highlighted so the readers will be informed about which country it is.

Overall, every graph has a tooltip function to provide the exact number and some context that readers will need to know. Most of the graphs have animations to show the changes in the trend of smoking across countries and years, or can be filtered by year by using the filter sliders. Our major interactions are selecting, clicking, and hovering over.

## Usability Findings

### Participant profile:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **Gender** | **Age** | **Career Background** | **Smoked**  **(Y/N)** | **Smoking**  **(Y/N)** |
| Participant 1 | female | In her 90s | No | Y | Y |
| Participant 2 | female | 22 | No | N | N |
| Participant 3 | male | 27 | PR | Y | Y |
| Participant 4 | male | 21 | No | Y | N |
| Participant 5 | male | 32 | Data visualization | N | N |

### Usability test findings:

**\*** are comments on the demo visualization

**Participant 1**

1. \*Didn’t know how to stop the animations
2. \*Didn’t notice the slider for graph and only click to explore
3. \*Suggested the filter should be right next to the corresponding graph
4. \*Not interested in the topic and keep neutral opinion
5. \*Tableau Online has time lag which makes the experience very unpleasant
6. \*Prefered more animations

**Participant 2**

1. \*Didn't understand what prevalence means
2. \*The grey color on the lines are not obvious on black grids
3. \*Suggested to add more colors to the background
4. \*Suggested to make the text bold
5. Animation control is too small to notice
6. Too many colors on the prevalence summary chart.

**Participant 3**

1. \*Color coding the bubble graph for total consumption is misleading. It might imply relations between countries.
2. \*Caught a mistake about the numbers on the y-axis for prevalence. It was 0-1000 for a variable in percentage.
3. \*It’s confusing what prevalence means

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Continue tests with revised visualization)

**Participant 4**

Tab 1:

1. Don’t know how to switch to the next page

Tab 2:

1. Don’t know what smoking prevalence means
2. Color for total consumption is confusing: color doesn’t indicate meaning or relation
3. Not sure which year it is on (“latest”) on total consumption map
4. Not sure what unit the line chart is in
5. Too much data in the treemap for number of smokers

Tab 3:

1. Don’t know how to clear after selecting items
2. Suggest to make the data exploration process from shallow to deep
3. Suggest to put the control of a graph at the bottom of the graph and the main control on top

Tab 4:

1. Not sure what unit prevalence is in
2. Suggest to relate to data other than prevalence, such as showing population in tooltips
3. Not necessary to use different color to represent countries
4. Want to select a country but cannot select from the legend

**Participant 5**

1. No unit for prevalence
2. Don’t quite understand the abbreviations for countries
3. Suggest to put the control at the bottom of the graphs
4. Suggested use pie chart for the number of smokers, otherwise the treemap looks a bit dazzled.
5. Suggest put control below the graph

## Summary of Results/recommendations

### **Summary of Results**

The ultimate goal of our team is to design an interactive visualization that enables readers to explore smoking-related data, in order to get some useful insights. Our guiding question, “how has the trend of smoking changed over the years from 1980 to 2012”, is only a starter. We want to leave the visualization as open-ended as possible in order to leave more room for readers to ponder. We hope to find out the advantages and flaws of our current visualization. After conducting five usability tests and consulting a professional in the field of data visualization, we have received many valuable comments and suggestions. We based our revisions on the feedback. The following is a summary of the results:

1. The positions of the filters make it difficult for the readers to tell which filter controls which graph. Some filters only impact one specific graph, while the others work like a master control that impacts every graph in the tab. Based on these comments, we now make the master control at the very top and the single filters at the bottom of and very close to each corresponding graph.
2. Most of the participants had trouble understanding what “prevalence” means. To solve this problem, we added percentage signs (%) on the graph where needed and added “xxx in total population” in tooltips to show that prevalence means a percentage in the total population.
3. Another thing that makes readers hard to interpret the graphs is the lack of units for axes. They didn’t know if the number is in the thousands, millions, or percent. Thus, we added units on every axis.
4. We were suggested to layout the graphs from shallow to deep analysis so that the readers will more easily understand the data. We took the advice and went from the overall analysis to a more specific analysis of each facet.
5. Lastly, some participants gave feedback about the color-coding we used. They said that the colors imply some relations between countries that are represented in similar colors. We tried using different color coding techniques, such as spectrum colors, but spectrum colors make it seem like countries in similar colors are somehow related. We eventually decided to keep using color-coding because it helps to distinguish one bar from the other in the graph. We intentionally used very different colors so that the readers would not relate one country to another.

### **Recommendations**

The purpose of our visualization is data exploration. Instead of providing findings to readers, we want readers to use our visualization as a tool to discover some trends and the potential messages these trends are trying to deliver. There are some insights we gathered that we didn’t expect to find. For instance, We can easily see a clear increase in the total consumption of cigarettes from 1980 to 2012, but this is not because the global total consumption of cigarettes has increased. It’s because of the huge increase in China that has significantly pulled up the average total consumption. If we rule out China, the increase becomes less dramatic. This might imply a factor of increasing alternatives for cigarettes, such as vaping and marijuana. Thus, it recommends a direction for further research. Secondly, in most of the countries, male smokers are much more than female smokers. However, in some European countries, such as Sweden, there are more female smokers than male smokers. This could be due to many reasons, such as a shift in culture, that is worth researching.

During our decision process, we have made some tradeoffs when choosing what data visualization techniques to use. Firstly, we originally used a map to represent the global total consumption because the map lays out the countries, so it is easier to see on the map. However, we then added another line graph for total consumption because the map lacks the ability to show the changes in the trends of smoking over the 33 years, and also got rid of the year filter on the map as it doesn’t really help to show the difference across the year. Secondly, we originally used a treemap to show the number of smokers for all countries, but the treemap was divided into so many meaningless small squares. Also, the total area of the treemap changes corresponding to the year it is in, which makes no sense. Thus, we changed it to pie charts. At first, the total area also changed, but we tackled this problem by changing the scaling from size to angle. Thirdly, even if we have received some feedback about the color-coding technique on the prevalence bar graph, we have still decided to keep it because we find it easier to distinguish between countries by using very different colors (colors that are not in the same spectrum). Lastly, there are some limitations with Tableau. Its page section, text, image, and position aren’t responsive, so the “automatic” size doesn’t work well. Even if we have tried many different size settings for the overall visualization, there still isn’t a perfect size that fits all screens.

We did data wrangling in Excel. Most of the data was easy to handle, except for the prevalence dataset. Due to its special characteristics of having too many factors and being in separate files for female, male, and both genders, we took very long to combine the three datasets into one comprehensive version.

Our overall recommendation based on the findings from these datasets is that in order to promote a healthy lifestyle, the governments do not only need to look into increasing taxes for cigarettes, but also pay more attention to the alternatives of cigarettes that also bring equivalent harm. Moreover, the government can have more specific regulations by analyzing the data. For example, data in some countries suggests paying more attention to the increasing smoking prevalence among teenagers, while some other data suggests that the increase in average cigarette consumption per smoker is a more pressing issue. Speaking from the data visualization perspective, it’s important to understand the characteristics of each chart type so that we can best select the ones the better answer our questions. We should always keep in mind what each factor, such as size and color, means so that we don’t make redundant design decisions. Lastly, color coding might make visualization look fancy, but we should always have a solid reason, other than polishing the visualization, to use the color-coding technique.

## Appendices

#### Pre-test Interview Questionnaire

1. What is your name? Age? Career background?
2. What is your nationality? Where do you live now?
3. Do you currently smoke or have you smoked before?
   1. If you smoked but quit, why did you quit?
4. Have you purchased cigarettes before, either for yourself or others?
5. What do you think about the trend of smoking today compared to 40 years ago?
6. Introduction to our interactive visualization on the smoking trend.
7. Ask the participant to think out loud.

#### Usability Test Questionnaire

**Visualization Orientation and Overview**

1. This is the intro page of our visualization.
   1. Can you tell what the bubbles represent?
   2. Why are some bigger than others?
   3. Do you have an idea of what we want to show you after you read this page?
   4. Do you know how to switch to the next page?
2. This is an overview of the topic.
   1. Where do you think you should start your exploration?
   2. Look at / scan the “World Cigarette Consumption” chart and explain what it shows. You are welcome to **click** on anything in the chart.
   3. Look at the three line charts below and try to explain what you see. Just hover over and think.
   4. Look at / scan the “Consumption Prevalence” and explain what it shows. You can **click on** anything in the chart and explain what it shows.

**Tasks**

1. Can you figure out the country with the largest amount of cigarette consumption on the world map? How do you know?
2. Select a country to explore as you like.
3. Can you figure out how to see the number of smokers in a certain country in 1994? What do the different colors mean? Can you show me how you found that information?
4. Can you figure out how the prevalence of smoking has changed over the years? Can you show me how you found that information?
5. Do you know how to make the animation stop?
6. Can you figure out how prevalent smoking is in 2005? Can you show me how you found that information?
7. If you think you have finished this page, you can move on now.
   1. Where do you think you should start your exploration?
   2. Look at / scan the “World Cigarette Smoking Prevalence” chart and explain what it shows. You are welcome to **click** on anything in the chart.
   3. Look at / scan the “Smoking Prevalence of All Ages” chart and explain what it shows. You are welcome to **click** on anything in the chart.
8. If you think you have finished this page, you can move on now.
   1. Look at / scan the chart and explain what it shows. You are welcome to **click** on anything in the chart.
   2. Why do some countries have larger areas than others?
   3. Can you figure out how the number of smokers has changed in different countries over the years?

**Which more makes sense?**

1. To make your expected chart by setting up the filters in a certain area such as at the top right corner of the visualization or to filter one graph by graph, which do you prefer?
2. What do you think of the order of each page? What do you expect to see one by one?
3. Can you understand what the visualization is about at the first sight of the captions above?
4. Is this visualization readable? What are some things you like/dislike about it?

**Closing**

Thanks to the test users for their participation. Ask them if they have any more feedback or questions.