CM2020-Agile Software Projects

Questionnaire Data Analysis, Group 8, Dec. 2021

In [2]:

```
!pip install matplotlib  # install the matplotlib, so that we can plot statistical graphs.
```

```
Requirement already satisfied: matplotlib in /opt/conda/lib/python3.7/site-packages
(3.2.1)
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.7/site-package
s (from matplotlib) (0.10.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /opt/cond
a/lib/python3.7/site-packages (from matplotlib) (2.4.7)
Requirement already satisfied: kiwisolver>=1.0.1 in /opt/conda/lib/python3.7/site-pa
ckages (from matplotlib) (1.2.0)
Requirement already satisfied: numpy>=1.11 in /opt/conda/lib/python3.7/site-packages
(from matplotlib) (1.18.4)
Requirement already satisfied: python-dateutil>=2.1 in /opt/conda/lib/python3.7/site
-packages (from matplotlib) (2.8.1)
Requirement already satisfied: six in /opt/conda/lib/python3.7/site-packages (from c
ycler>=0.10->matplotlib) (1.14.0)
WARNING: You are using pip version 20.2.3; however, version 21.3.1 is available.
You should consider upgrading via the '/opt/conda/bin/python3 -m pip install --upgra
de pip' command.
```

In [3]:

```
import matplotlib.pyplot as plt  # shorten the name so that it's simple to reuse
import numpy as np   # NumPy library operated numerical calculations, especially the matrix operati
import pandas as pd  # Pandas is a library based on NumPy. Pandas focuses on statistical analysis.
```

In [4]:

```
# In order to make the plots more beautiful:
import matplotlib as mpl  # import the plot drawing library and shorten the name
mpl.style.use('ggplot')  # adjusts the style to emulate ggplot (a popular plotting package for R)
```

In [5]:

```
import seaborn as sns  # statistical data visualisation tool which can draw high-level plots
```

In [6]:

```
# the below file is my dataset:
# the file is a combined result, with both Typeform results and Chinese results.
file = r'Combined_results.xlsx' # set the mode as 'only read', as we don't change the file
df = pd.read_excel(file) # 'df' is our data set's name
# df = pd.read_excel(file, encoding='gb18030') # I put an encoding since there was a Unicode Decode
df.head(23)  # show the first 23 rows of the data set
```

Out[6]:

	#	gender	from	job	Do you know about Tower Defense Games? (你知道塔防游戏吗？)	How often do you play Tower Defense Games? (多久玩一次塔防游戏？)	Of the tower defence games you like, what game(s) did you like? (塔防游戏中，你最喜欢哪个？)	diff w you l t def gan be? 喜欢 游 度
0	ywtewt987uyfj1m1ywtew0t1y23dj8s6	Female (女)	Europe (欧洲)	Student (学生)	Yes, it's just building towers and then defens...	Seldom (从不)	NaN	Toug fair (但

In [84]:

```
df.shape  # how many rows, how many columns.
```

Out[84]:

(17, 26)

In [85]:

```
# Now we get to know the basic statistics for each column, using 'describe()' function:
df.describe()
```

Out[85]:

	#	gender	from	job	Do you know about Tower Defense Games? (你知道塔防游戏吗？)	How often do you play Tower Defense Games? (多久玩一次塔防游戏？)	Of the tower defence games you like, what game(s) did you like? (塔防游戏中，你最喜欢哪个？)	How difficult would you like a tower defence game to be? (你喜欢哪种游戏难度？)	Cute & cartoon (卡通可爱)	War-related (战争相关)	...
count	17	17	17	17	17	17	14	17	9	8	...
unique	17	2	3	8	3	4	14	3	1	1	...
top	Yilong	Male (男)	Asia (亚洲)	Student (学生)	Yes, it's just building towers and then defens...	Seldom (从不)	Haven't played it. 没玩过	Tough but fair (难，但有道理)	Cute & cartoon (卡通可爱)	War-related (战争相关)	...
freq	1	9	11	4	9	11	1	15	9	8	...

4 rows × 26 columns

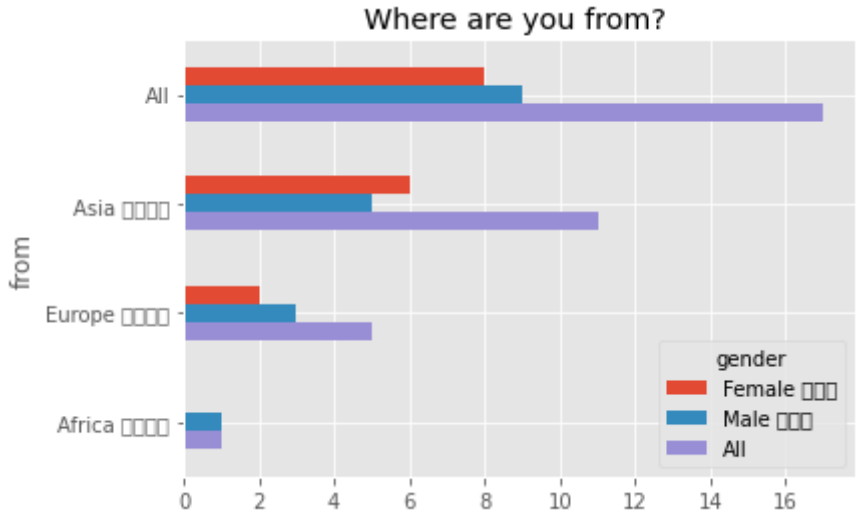
Data set descriptions:

- **Sample size: 17. Chinese samples: 7; world-wide samples: 10.**
- Since there are question amount restrictions in the Typeform website, we have to limit our questions and delete others. There are 10 questions in Typeform questionnaire in total, and this form is prepared for the world-wide group. Also, there is a sample size limit for the Typeform, so we only have 10 samples.
- Since Xiaoyun Yu is just based in mainland China, she can collect opinions from the people surrounding her. She believes that in this way we can obtain more opoinions, so she has made another questionnaire

form using Tencent form collection method. There is no restrictions for question amount or sample size, so for the questionnaire's China version, we have asked almost all of our concentrated questions there.
- Finally, we have collected all the results into one data file - 'Combined_results.xlsx', where we only list the common 10 questions for both versions.

In [86]:

```
table = pd.crosstab(df['from'],
                    df['gender'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='All', ascending=False)
# table.drop('All', inplace=True)

ax = table.plot.barh()
ax.invert_yaxis()
ax.set_title('Where are you from? ')
plt.show()
```

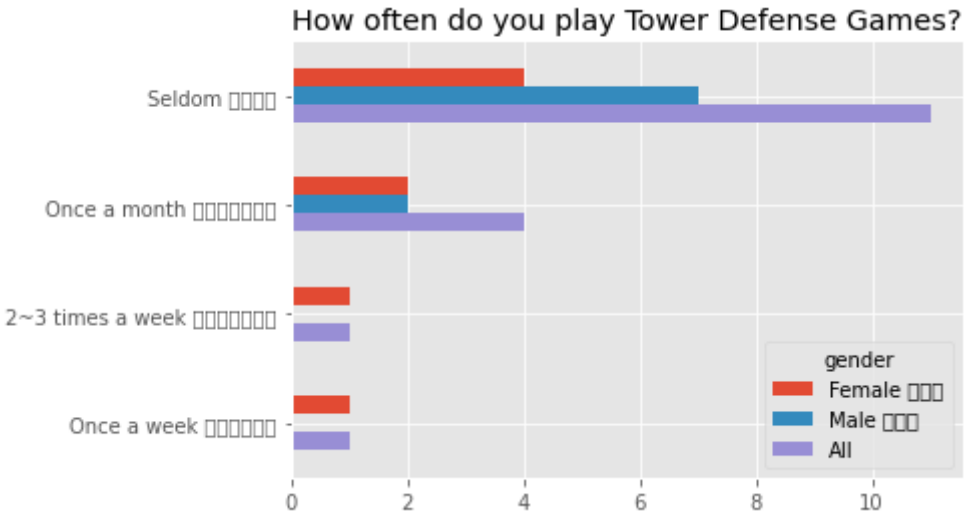


Insights:

- There are 8 female guests, and 9 male guests in total. They are almost equal, which is quite good for us to analyze gender requests.
- **Most of our guests are from Asia (11 people).** 7 are from mainland China that's certain, and the other 4 may come from other areas.
- We have 5 Europe guests, and 1 Africa guest. These samples may help us to understand the Europe market, which perhaps is different from Asians' thoughts.
- Now we have questions: Do Asians have the same preference or different preferences from the European or African people? Can we combine these preferences?
- We need to decide our market range based on the location question - Do we need to proporgate our product only to Europe-America market? Or do we want only the Asia market? Or do we have bigger ambition to march to world-wide market (Can we combine all the interests)? What's our focus?

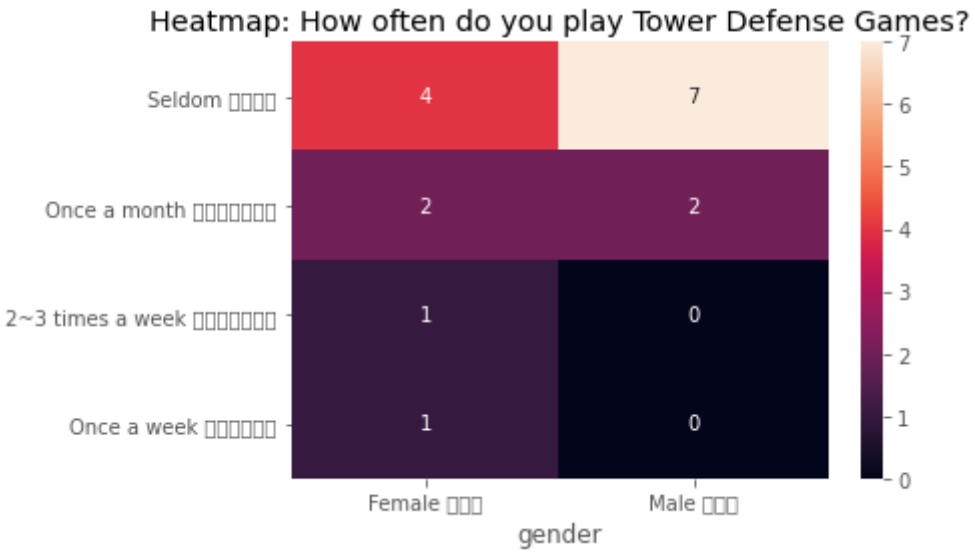
In [56]:

```
table = pd.crosstab(df['How often do you play Tower Defense Games? （多久玩一次塔防游戏？）'],
                    df['gender'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='All', ascending=False)
table.drop('All', inplace=True)

ax = table.plot.barh()
ax.invert_yaxis()
ax.set_title('How often do you play Tower Defense Games?')
ax.set_ylabel(' ')
plt.show()
```



In [55]:

```
# Heat map:
table = pd.crosstab(df['How often do you play Tower Defense Games? （多久玩一次塔防游戏？）'],
                    df['gender'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='Female （女）', ascending=False)
table.drop('All', axis='columns', inplace=True)
table.drop('All', inplace=True)

ax = sns.heatmap(table, annot=True)
ax.set_title('Heatmap: How often do you play Tower Defense Games?')
ax.set_ylabel(' ')
plt.show()
```

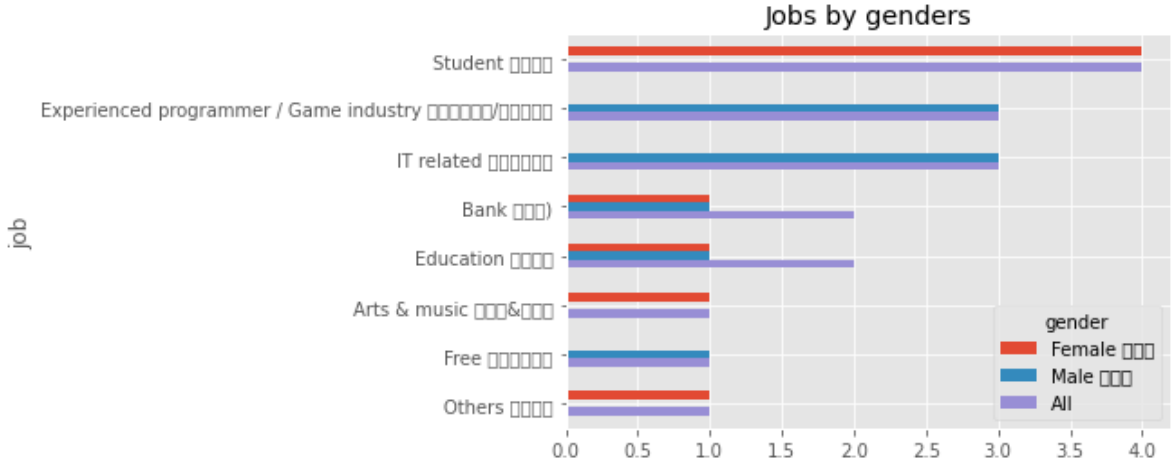


Insights:

- **Most of the people in our sample, don't play TD games at all or just play them quite few times.**
- The ones who play TD games more oftern, are females.

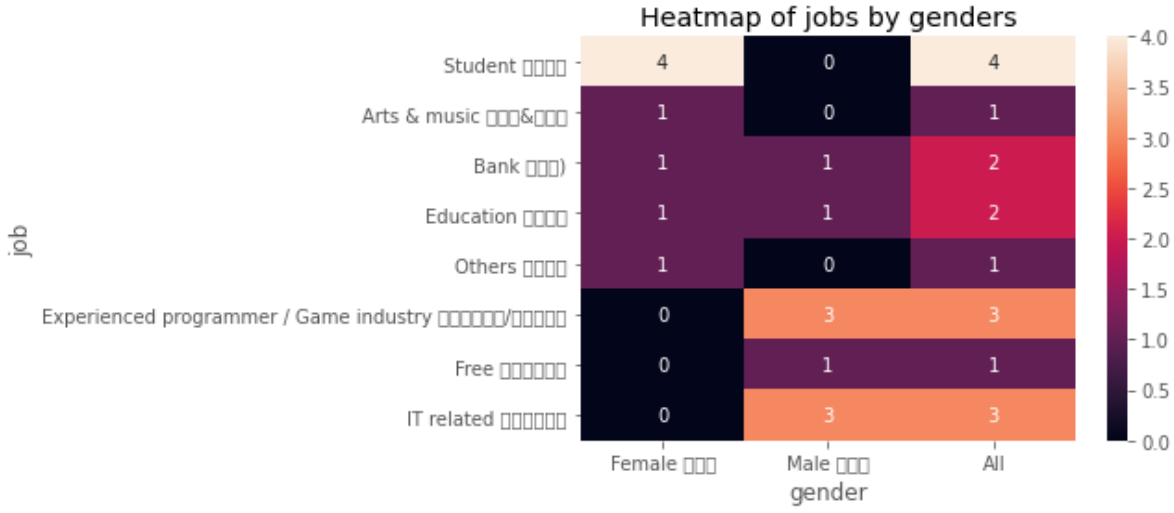
In [53]:

```
table = pd.crosstab(df['job'],
                    df['gender'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='All', ascending=False)
table.drop('All', inplace=True)

ax = table.plot.barh()
ax.invert_yaxis()
ax.set_title('Jobs by genders')
plt.show()
```



In [51]:

```
# Heat map:
table = pd.crosstab(df['job'],
                    df['gender'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='Female （女）', ascending=False)
# table.drop('All', axis='columns', inplace=True)
table.drop('All', inplace=True)

ax = sns.heatmap(table, annot=True)
ax.set_title('Heatmap of jobs by genders')
plt.show()
```



Insights:

- We settle this job question, in order to: (1) figure out whether our data sample has **any bias with regard to occupations**. For example, do most of the answers come from programmers? If so, then there would be bias about their theme preference, topic preference, play duration and frequency, and so on. We need to emphasize our goal - we wish our product can be used by common people, and, we wish from our questionnaire, we can hear diversified voices so that later on we can settle down our target user groups.
- (2) If there are experienced programmers, then no problem, we have this honour to hear from them about **any advice** on our product, as we are all inexperienced to make TD game.
- There are many students. And of course, our main target group would be college students who are free for time and most possible to play video games.
- There are 3 experienced programmers, and 3 IT-related people, and 2 people working in educational positions. We need to be careful about any bias this may bring for us, and we indeed look forwared to their suggestions.

In [52]:

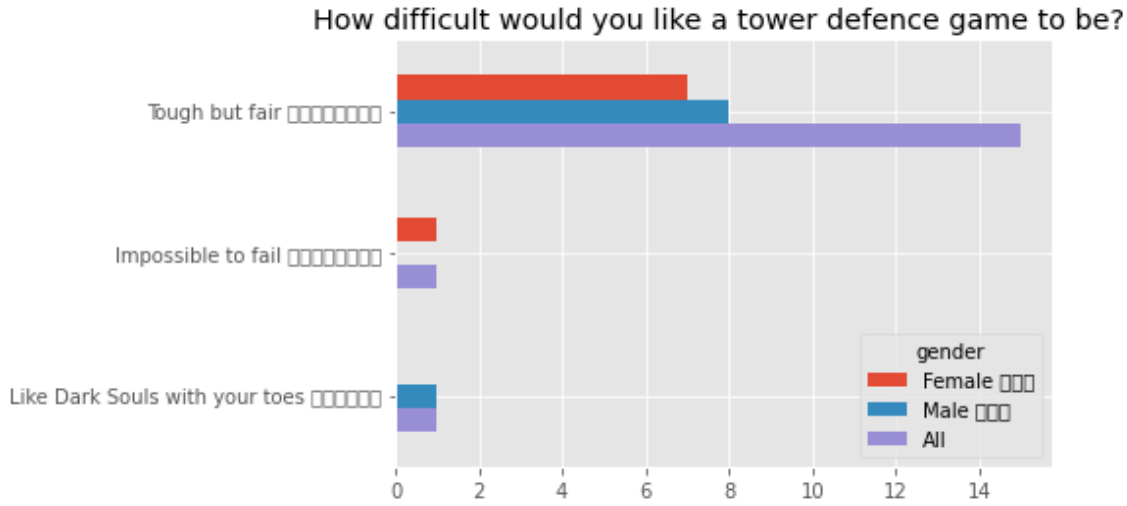
```
table = pd.crosstab(df['Do you know about Tower Defense Games? （你知道塔防游戏吗？）'],
                    df['gender'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='All', ascending=False)
table.drop('All', inplace=True)

ax = table.plot.barh()
ax.invert_yaxis()
ax.set_title('Do you know about Tower Defense Games? ')
ax.set_ylabel(' ')
plt.show()
```

Insights:

- Interestingly, although many people seldom play TD games, still most of the people claim that they know what TD games are. Perhaps it's because the misleading of our choices, and people always don't want to admit that they have unknown knowledge about some simple concepts.
- There are 4 people state that they don't know about TD games at all. It means, for common people, there is **big possibility that our TD game would be their first TD game**. So we should make more tutorials, and, make more attrctive elements.

In [66]:

```
table = pd.crosstab(df['How difficult would you like a tower defence game to be? （你喜欢哪种游戏难度
                    df['gender'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='All', ascending=False)
table.drop('All', inplace=True)

ax = table.plot.barh()
ax.invert_yaxis()
ax.set_title('How difficult would you like a tower defence game to be? ')
ax.set_ylabel(' ')
plt.show()
```



Insights:

- Most of the people hope to play games which are **a little bit difficult but with some sense**.
- One female hopes to play a game that never fails. That means, she hopes to relax as much as she could.
- One male hopes to challenge his max ability.

In [59]:

```
cute = df['Cute & cartoon （卡通可爱）'].count()
cute
```

Out[59]:

9

In [60]:

```
war = df['War-related （战争相关）'].count()
war
```

Out[60]:

8

In [61]:

```
zombie = df['Zombie （僵尸）'].count()
zombie
```

Out[61]:

1

In [62]:

```
educational = df['Educational （教育）'].count()
educational
```

Out[62]:

4

In [63]:

```
abstract = df['Abstract （抽象）'].count()
abstract
```

Out[63]:

4

In [65]:

```
others = df['Others （其他）'].count()
others
```

Out[65]:

1

In [67]:

```
theme_dictionary = {'cute': cute, 'war': war, 'zombie': zombie, 'educational': educational,
                    'abstract': abstract, 'others': others}
theme_dictionary
```

Out[67]:

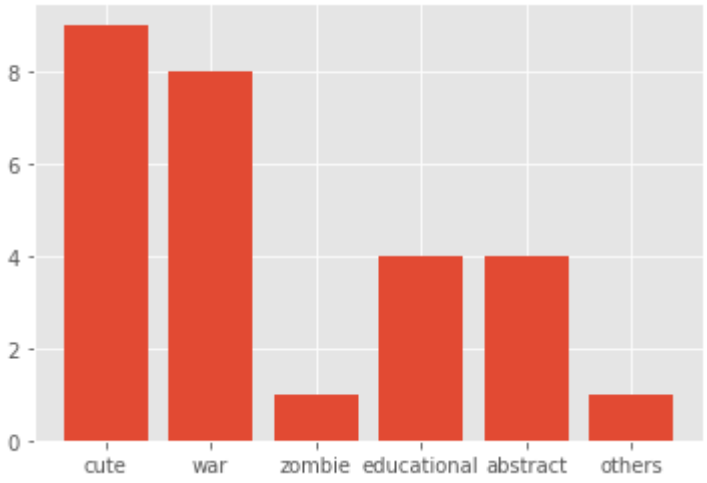
```
{'cute': 9,
 'war': 8,
 'zombie': 1,
 'educational': 4,
 'abstract': 4,
 'others': 1}
```

In [93]:

```
# bar plot for themes
plt.bar(theme_dictionary.keys(), theme_dictionary.values(), 0.8)
```

Out[93]:

<BarContainer object of 6 artists>



Insights:

- **'Cute & cartoon' receives the highest** vote rate. Some people who choose 'war-related' also have chosen 'cute & cartoon'. It means, no matter the theme is, we should make our game cartoon-like, instead of making it similar to real people or device.
- We have doubted whether 'cute & cartoon' is only preferred by Asians. But in fact, if we look into the answer chart, we can see that many European people and even males choose 'cute & cartoon'. It means, this choice is people's choice.
- The choice with silver gold is **'war-related'**. Both Asians and people from other places choose this choice. It means, we should make our game engaged with war-related elements - rivalry, weapons, fire, battlefield, etc. We should discover the attractive elements of the battlefield, then cartoonize them.
- For 'educational' the theme we care the most, we find that there are indeed 4 people support it. Of course our guest who works in education area has chosen this choice, and there are still other people would like to see educational elements.

In [74]:

```
maths = df['Maths （数学）'].count()
maths
```

Out[74]:

9

In [75]:

```
english = df['English （英语）'].count()
english
```

Out[75]:

3

In [77]:

```
other_languages = df['Other languages （其他语种）'].count()
other_languages
```

Out[77]:

6

In [78]:

```
science = df['Science （科学）'].count()
science
```

Out[78]:

6

In [79]:

```
history = df['History （历史）'].count()
history
```

Out[79]:

7

In [81]:

```
literature = df['Literature （文学）'].count()
literature
```

Out[81]:

3

In [88]:

```
music = df['Music （音乐）'].count()
music
```

Out[88]:

6

In [87]:

```
non_edu = df['non-educational theme （不喜欢教育主题）'].count()
non_edu
```

Out[87]:

```
4
```

In [89]:

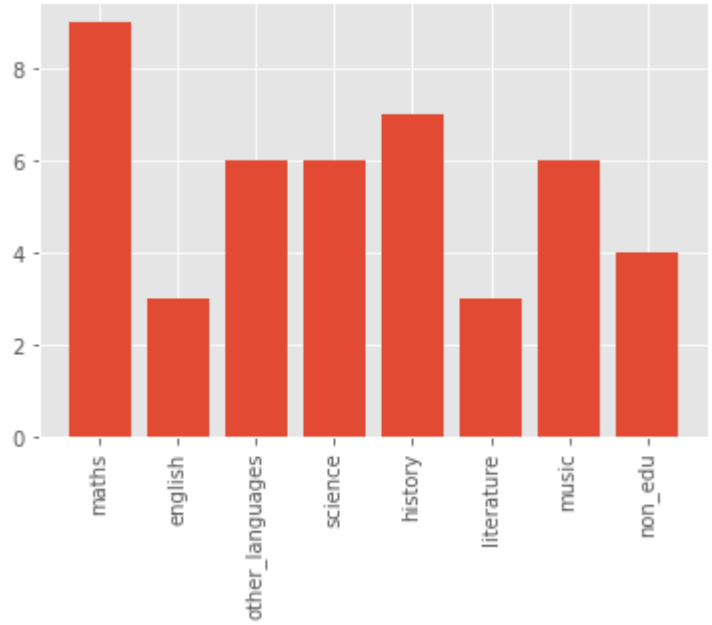
```
subject_dictionary = {'maths': maths, 'english': english, 'other_languages': other_languages,
                      'science': science, 'history': history, 'literature': literature,
                      'music': music, 'non_edu': non_edu}
subject_dictionary
```

Out[89]:

```
{'maths': 9,
 'english': 3,
 'other_languages': 6,
 'science': 6,
 'history': 7,
 'literature': 3,
 'music': 6,
 'non_edu': 4}
```

In [91]:

```
plt.bar(subject_dictionary.keys(), subject_dictionary.values(), 0.8)
plt.xticks(rotation=90)
plt.show()
```



Insights:

- There are 17 people in our sample, and there are 4 people have chosen 'I don't like educational theme'. According to this result, we should be **very careful about how to implement any educational elements**. Are they boring? Are they serious? Are they tough? These could all lead to the loss of market.

- In our result, **'maths' wins**. Perhaps some of the reason is that there are IT-related people and many programmers. Anyway, we should be confident if we hope to make a maths-related game.
- Since I've taken samples in mainland China, people seem to be lack of interests in learning English.
- People would like to learn history. We can consider whether we will make a game with a background taken from real history.
- People like science topic.
- There are indeed many more subjects and topics not included here.

Now let's check the additional questions asked in Chinese version questionnaire:

In [7]:

```
# the below file is my dataset:
# the file is a combined result, with both Typeform results and Chinese results.
origin = r'Chinese_responses.xlsx' # set the mode as 'only read', as we don't change the file
dfc = pd.read_excel(origin) # 'df' is our data set's name
# df = pd.read_excel(file, encoding='gb18030') # I put an encoding since there was a Unicode Decode
dfc
```

Out[7]:

	People 提交者（自动）	Submission Time 提交时间（自动）	Gender 你的性别？（必填）	Where are you from? 你来自世界上的哪里？（必填）	Job 你的工作？选一个最贴切的（必填）	Do you know TD games? 你知道塔防游戏吗？（必填）	How often do you play TD games? 你多久玩一次塔防游戏？（必填）	Of the tower defence games you like, what game(s) did you like? 塔防游戏中，你最喜欢哪个？	d y a d ga
0	刘伊琳	2021-11-30 17:00:48	Female (女)	Asia (亚洲)	Bank (银行)	Somewhat, I've played Plants vs. Zombies, but ...	Once a month (一个月一次)	NaN	
1	送快递的狐狸喵	2021-11-30 17:20:20	Male (男)	Asia (亚洲)	IT related (信息技术)	Yes, it's just building towers and then defens...	Seldom (从不)	In Wechat small program, there is a sth. Calle...	
2	利剑	2021-11-30 17:32:54	Male (男)	Asia (亚洲)	Bank (银行)	No, I've never played it before. (不知道，没玩过)	Seldom (从不)	Don't know. 不知道	
3	TuringSnowy	2021-11-30 18:06:49	Male (男)	Asia (亚洲)	Education (教育)	No, I've never played it before. (不知道，没玩过)	Seldom (从不)	Haven't played it. 没玩过	

4	七柱🏯篝火🍥	2021-11-30 18:43:45	Female (女)	Asia (亚洲)	Student (学生)	Yes, it's just building towers and then defens...	一周一次	Defending Radish. 保卫萝卜
5	Kaikai	2021-12-01 03:19:12	Female (女)	Asia (亚洲)	Arts & music (艺术&音乐)	Somewhat, I've played Plants vs. Zombies, but ...	Once a month (一个月一次)	Tank battles. 几百年前小霸王游戏机里的坦克大战
6	Yilong	2021-12-01 08:12:44	Male (男)	Asia (亚洲)	Free （自由职业）	Somewhat, I've played Plants vs. Zombies, but ...	Once a month (一个月一次)	Montizuma's Anger. 蒙提祖玛之怒

In [8]:

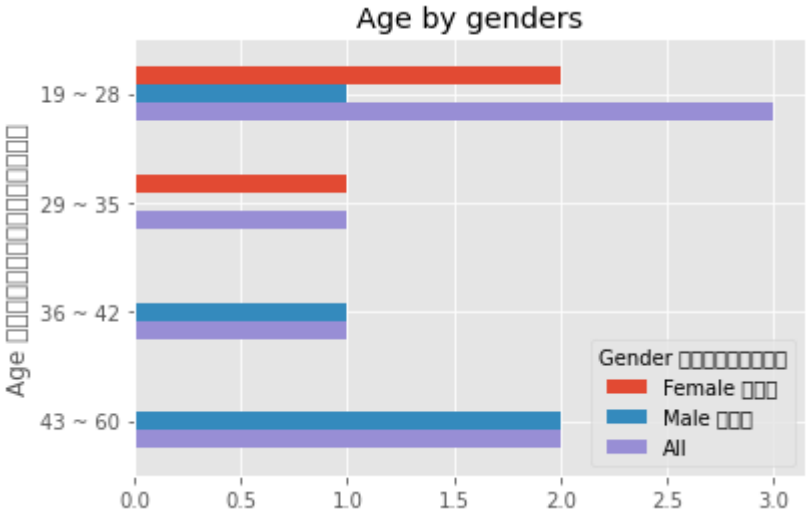
```
dfc.describe()
```

Out[8]:

	People 提交者（自动）	Submission Time 提交时间（自动）	Gender 你的性别？（必填）	Where are you from? 你来自世界上的哪里？（必填）	Job 你的工作？选一个最贴切的（必填）	Do you know TD games? 你知道塔防游戏吗？（必填）	How often do you play TD games? 你多久玩一次塔防游戏？（必填）	Of the tower defence games you like, what game(s) did you like? 塔防游戏中，你最喜欢哪个？	How difficult would you like a tower defence game to be? 你喜欢哪种游戏难度？（必填）	t de d pr 你 哪 型 的 游
count	7	7	7	7	7	7	7	6	7	
unique	7	7	2	1	6	3	3	6	2	
top	利剑	2021-12-01 03:19:12	Male (男)	Asia (亚洲)	Bank (银行)	Somewhat, I've played Plants vs. Zombies, but ...	Seldom (从不)	Tank battles. 几百年前小霸王游戏机里的坦克大战	Tough but fair (难，但有道理)	C ca (-
freq	1	1	4	7	2	3	3	1	6	
first	NaN	2021-11-30 17:00:48	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
last	NaN	2021-12-01 08:12:44	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

In [11]:

```
table = pd.crosstab(dfc['Age 你的年龄？我们会保密哒（必填）'],
                    dfc['Gender 你的性别？（必填）'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='Age 你的年龄？我们会保密哒（必填）', ascending=True)
table.drop('All', inplace=True)

ax = table.plot.barh()
ax.invert_yaxis()
ax.set_title('Age by genders')
plt.show()
```

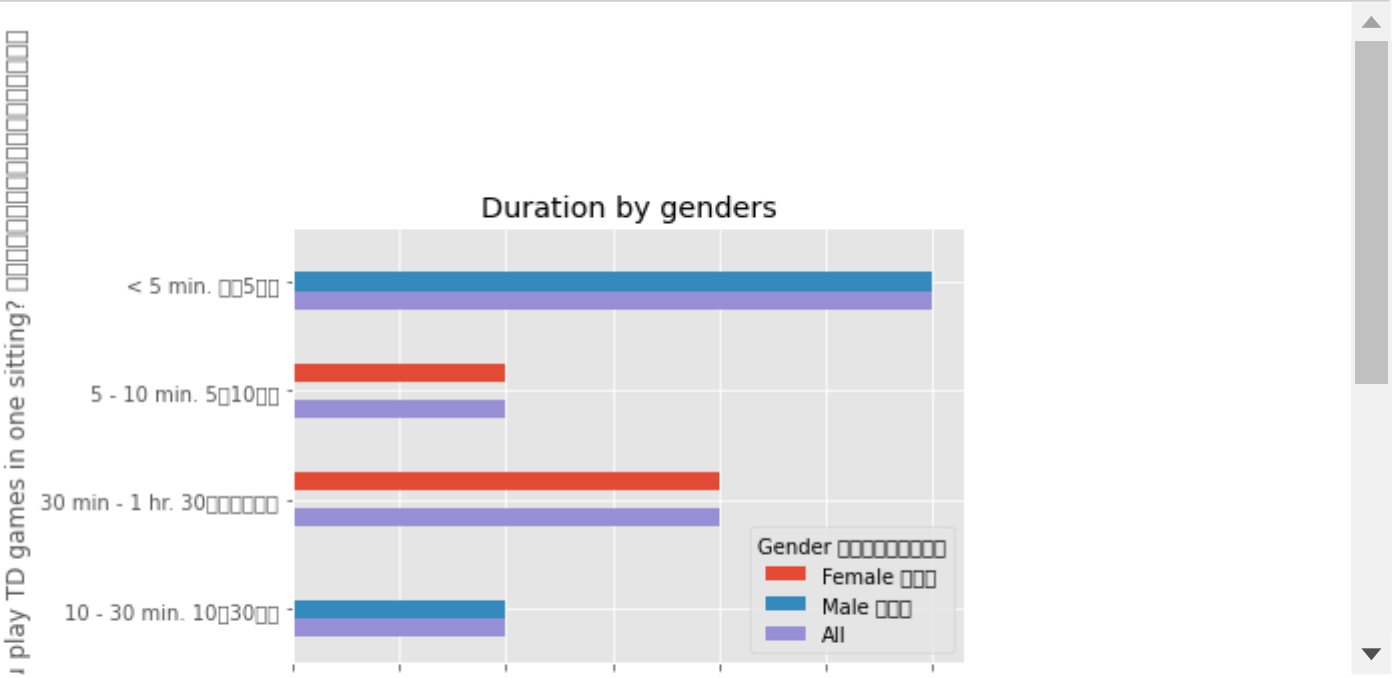


Insights:

- Top 1 group people in our sample (mainland China version) is young people.
- We have also many samples from middle-aged people.
- Many sampled people here are young females, and middle-aged males.

In [14]:

```
table = pd.crosstab(dfc['How long do you play TD games in one sitting? 你每次玩塔防游戏一般玩多久?
                    dfc['Gender 你的性别? （必填）'], margins=True)
# table.drop('All', axis='rows', inplace=True)
table = table.sort_values(by='How long do you play TD games in one sitting? 你每次玩塔防游戏一般玩多
table.drop('All', inplace=True)

ax = table.plot.barh()
ax.invert_yaxis()
ax.set_title('Duration by genders')
plt.show()
```



Insights:

- Almost a half of the people here only play TD games < 5 min, which means they seldom play this type of game. We should consider how to attract these group of people.
- Surprisingly, there are 3 females play TD games > 30 min in one sitting. We should consider how to meet their needs in developing a game with continous stories. And, we should consider how to constrain their usage of screen for protection of their health, as we truly hope to educate people.

In [ ]:

In [ ]:

Suggestions from all answers collected:

- 'Ability to speed up attacks if there is a level system, so you dont have to wait for the attack and defense to play out. '
- 'Skin customization. '

- 'More levels. '
- 'More fair less grind.'
- 'Naming squads or troops ability. '
- 'Bigger freedom for enemy's movement. '
- 'More skills and magic. '
- 'Prefer less payment. '
- 'I am slow to accept new things. Don't update the types of tower too often. Don't like big games. '
- 'I hope Plants vs Zombies can allow users to design the map or site just like Tank Battles. '
- 'I hope the shortcut keys would be more user-friendly. '

Our (Chinese) users hope to see these elements:

- 'Plants, animals, foods. '
- 'I hope to see the "Pass" sign at each level. I hope to own many types of weapons at the beginning, and my lives are full. I like the beautiful and bright scene, rather than dark colour design. '
- 'Still eyesight to prevent from feeling dizzy. Simple and small games with a sense of design. '
- 'Design. '

In [ ]:

In [ ]:

In [ ]: