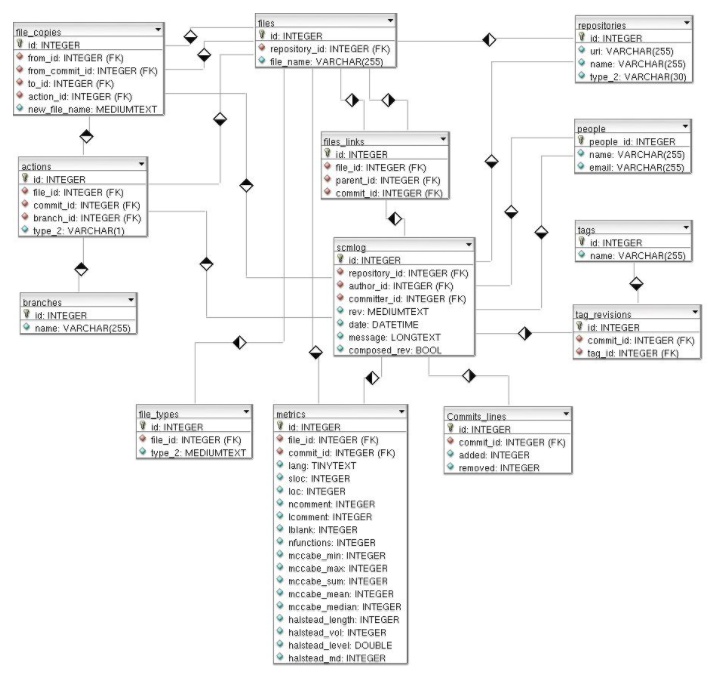
**Visual Exploration Exercise Report**

**D3.js git log analysis using Tableau**

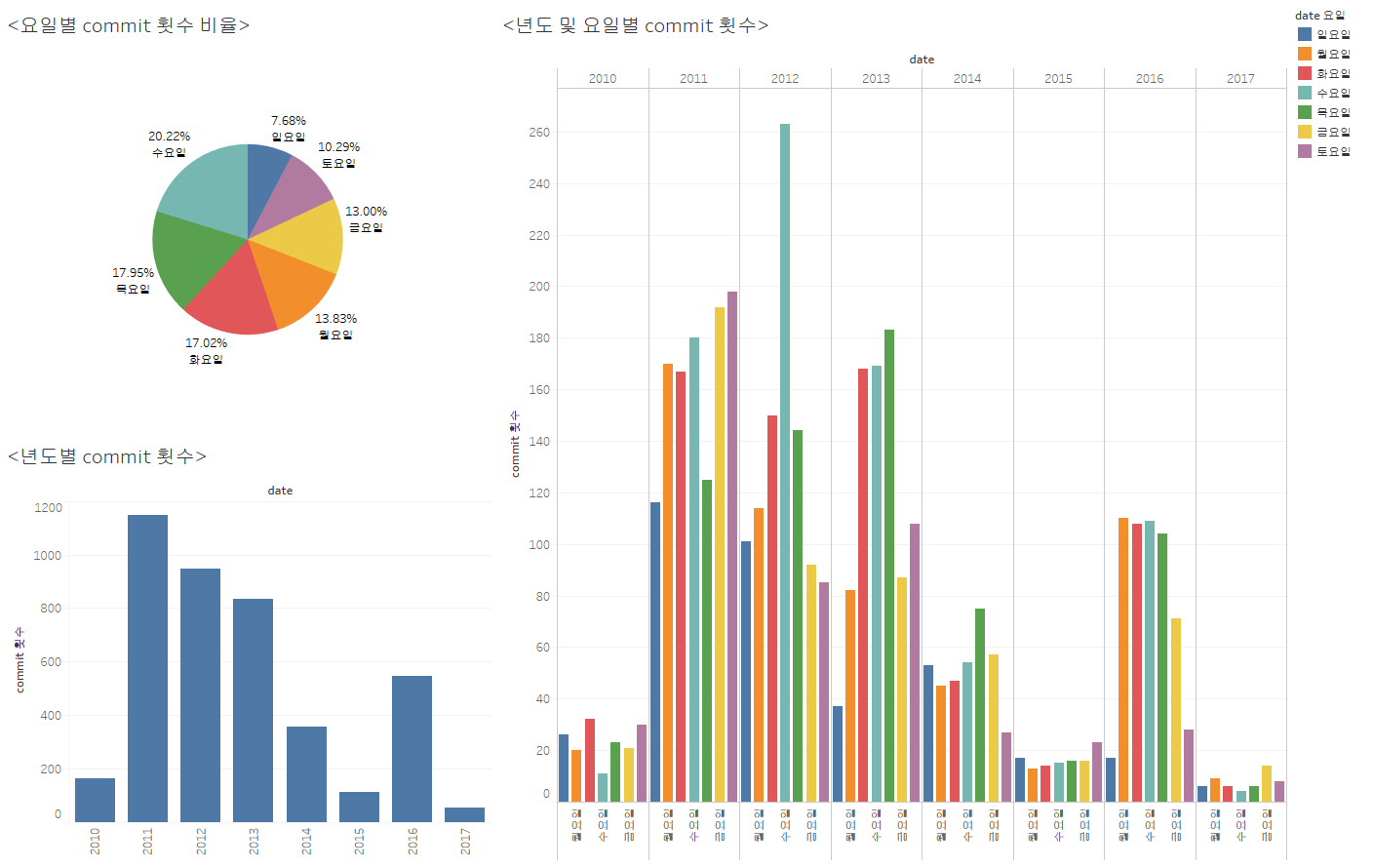
컴퓨터공학부

2017-24950

김이은

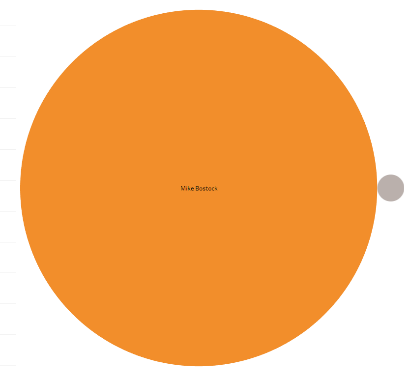
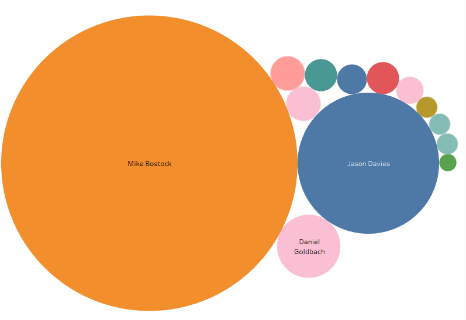
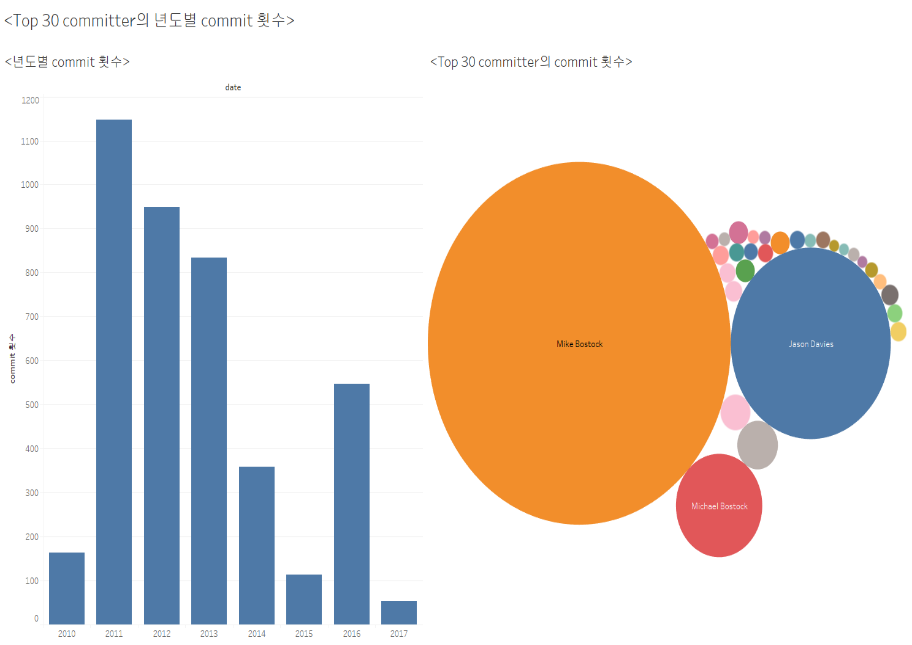
1. **Data / Infovis Tool for Analysis**  
   프로젝트 proposal에 언급한 내용을 토대로, 프로젝트에서는 주로 개발자의 SCM 및 ITS 관련 데이터를 다룰 계획이다. 이번 과제에서는 프로젝트와 관련된 데이터를 사용하기 위해서 d3.js git log 데이터에 대한 분석을 진행했다. D3.js git log data를 cvsanaly tool을 사용해 분석했고, dataset구조는 아래의 표와 같다. 각 dataset에 대해 csv file을 생성해 Tableau의 입력 데이터로 사용했다. Commit 수로는 총 데이터가 4166개이고, 10개의 csv file을 생성했다.
2. **Interesting Findings (insights or discoveries)**

* **scmlog dataset**  
  D3.js 개발이 시작된 2010년부터 현재(2017년)까지 각 연도 및 요일 별로 commit 횟수에 대해 분석했다. Figure1의 왼쪽 아래에 각 연도별 commit 횟수를 막대 그래프로 표현했다.   
    
  **Insight1)** 2011년에 release된 이후로 version update가 2011년 8월 23일, 2012년 12월 21일, 2016년 6월 28일로 총 3번 있었는데, Figure 1의 왼쪽 아래 막대 차트에서 단순히 해당 연도들의 commit 횟수를 살펴보면 전반적으로 commit 횟수가 많음을 발견할 수 있다.  
    
  **Insight2)** Figure 1의 왼쪽 위에서 전체적인 요일별 commit 횟수 비율을 살펴보면, 수요일이 20.22%로 가장 높은 비율을 차지하지만 목요일(17.95%), 화요일(17.02%)과 비교했을 때 큰 차이가 없음을 알 수 있다. 그리고 주말인 토, 일요일은 10% 이하로 가장 낮은 비율을 차지하는 것을 확인할 수 있다. 그러나, 대시보드의 오른쪽 막대그래프를 살펴보면 각 연도 내에서 수요일이 가장 commit 횟수가 높은 것이 아님을 알 수 있다. 근소한 차이로 수요일이 가장 높은 commit 횟수 비율을 차지하게 된 이유는 2012년의 수요일의 commit 수가 263개로 다른 요일들에 비해 압도적으로 높기 때문이다.

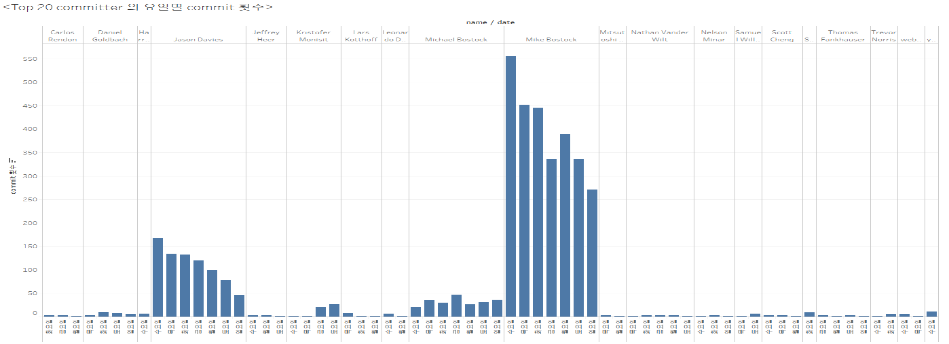


**Figure 1년도 및 요일별 commit 횟수**

* **scmlog / people datasets**  
  scmlog, people dataset을 이용해 committer 중 가장 많은 commit을 한 순서대로 상위 30명을 필터링 기능을 사용해 추출했다. 그리고 Tableau의 대시보드 기능을 사용해 전반적으로 d3.js 개발에 참여한 top 30 committers 중 각 연도별로 어떤 committers들이 활발했는지 분포를 살펴보았다. Filter 기능을 사용해 각 연도별로 활발하게 commit한 committer들을 Figure 2 오른쪽의 버블 차트로 표현했다.  
    
  **Insight3)** top committer 중에서 mike bostock와 Jason Davies가 매년 많은 횟수의 commit을 하고 있음을 알 수 있다. 그리고 Top 30 committers 에 대해 2010년경에는 major하게 참여하는 committer수가 3명 (Michael Bostock, Mike Bostock, Jeffrey Heer)인 반면에 2011년~2015년에는 약 10으로 증가했음을 알 수 있다. 하지만, 최근 2016, 2017년에는 major committer는 약 2명으로 감소했다.  
    
  **Insight4)** Insight3에서Mike Bostock이 압도적으로 commit 을 많이 했음을 알 수 있다. 그리고 scmlog dataset에서 수요일에 가장 많은 commit을 한다는 사실을 발견했는데, 이러한 사실이 Major committer Mike Bostock의 commit 요일이 수요일이기 때문에 나온 것인지 궁금증이 생겼다. 그래서 Top 20 committer에 대해 요일별 commit 횟수를 막대 차트를 통해 살펴보았다. Mike Bostock 이 수요일에 가장 많은 commit을 했고 절대적인 commit 횟수가 많기 때문에 전체 요일 commit 수에 큰 영향을 미치지만, 다른 committer들도 수요일이 가장 높고, 목요일, 화요일 순서임을 알 수 있다.



**Figure 2 Top 30 committer의 연도별 commit 횟수**



**Figure 3 Top 20 committer의 요일별 commit 횟수**

* **scmlog / tag revision / tag datasets**  
  **Insight5)** Tag dataset은 version history 정보를 포함하고 있다. Tag를 단 committers 은 총 4명으로 mike bostock, michael bostock, jason davies, square가 있다. 이들의 tag revision 비율을 Figure 4와 같이 트리맵 형태로 살펴보았다. 이를 통해 Version update에서도 mike bostock이 압도적인 참여했음을 알 수 있다.



**Figure 4 committers의 tag revision 비율**

1. **Infovis tool positive/negative aspects and suggestions for improvements**- spotfire에 비해 주관적으로 인터페이스가 단순하고 미적인 느낌을 준다. novice의 접근성 측면에서 활용도가 좋을 것 같다.  
   - 데이터를 살펴볼 때, 시트에서 나가서 원본데이터로 가서 살펴봐야 해서 불편했다. 차트와 나란히 놓고 비교해서 볼 수 있으면 좋을 것 같다.  
   - Color 군이 다양하지 않은 점이 좀 아쉬웠다. 직접 color를 선택할 수 있지만, 이는 Color 개수나 데이터의 특성에 적합한지 등의 고려해야할 요소들이 많아 어려움이 있다. 그래서 정해진 color 군 내에서 몇 가지의 color는 직접 선택해서 바꿀 수 있도록 flexibility를 제공하는 것도 좋을 것 같다.   
   - filter를 적용한 후 원하는 부분만 화면을 캡쳐할 때, filter가 고정된 박스가 테두리가 생기고 옆에 기능들이 보이면서 부가적인 정보들이 캡쳐 화면에 포함되는 것이 거슬렸다..