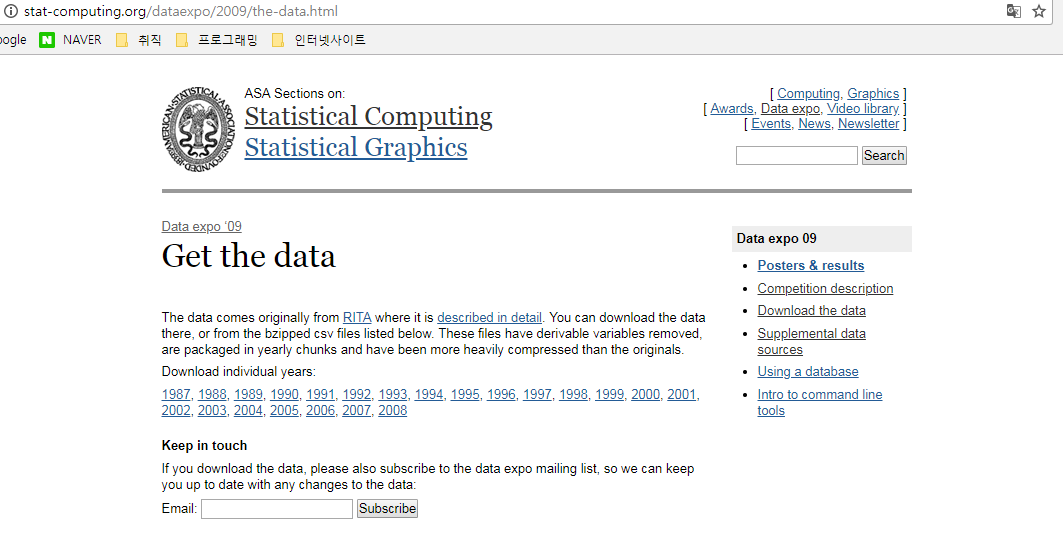
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **□ 수행평가 – 빅데이터 분석시스템 구축 과정** | | | | |
|  |  |  |  |  |
| **과정명** | | [4차산업 선도인력]커넥티드 카를 위한 IoT 전문가 과정(NCS) | | |
| **능력단위명** | | 분석용 데이터 탐색 2001010505\_15v1 | **훈련교사** | 이진만 |
| **능력단위요소명** | | 빅데이터 저장 계획 수립하기, 빅데이터 저장 모델 설계하기 빅데이터 저장 관리시스템 구성하기 빅데이터 적재하기, 빅데이터 운영하기 | | |
| **수행날짜** | | 2018.04.05 | 훈련생명 | 이태우 |
| **과제개요** | | | | |
| 1. 데이터를 수집 한다.(공공데이터, 서울시 공공데이터) 2. 데이터를 빅데이터 시스템(하둡 및 데이터베이스) 저장  3. R을 통해 데이터를 탐색 및 분석 한다 4. 데이터 분석 화면을 구현 한다. | | | | |

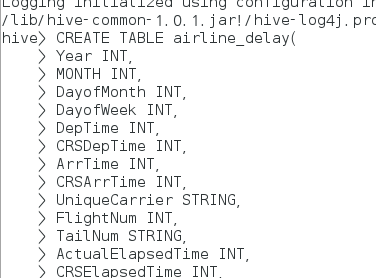
1. 데이터를 수집 한다.(공공데이터, 서울시 공공데이터)



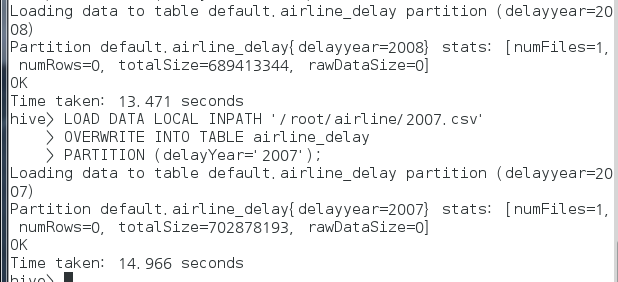
🡪미국 airline\_delay 2006~2008까지 데이터 수집

1. 데이터를 빅데이터 시스템(하둡 및 데이터베이스) 저장

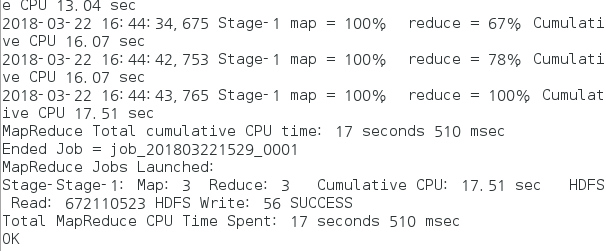
**태이블 생성**



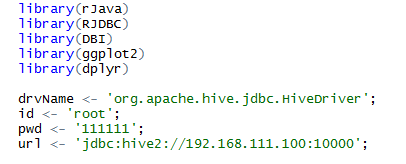
**DATA load 를 해준다**



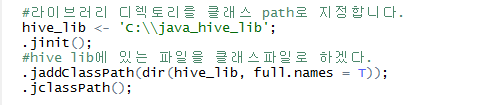
**Hive --service hiveserver2 를 실행시켜 연동을 시킨다**

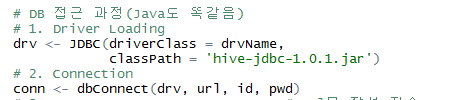


1. R을 통해 데이터를 탐색 및 분석 한다

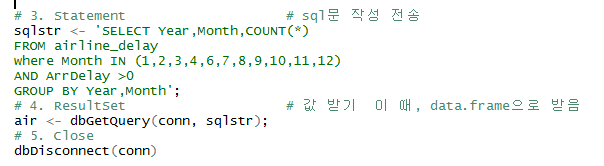


🡪library를 임포트해준다, drvName,id,pwd,url을 입력하여준다

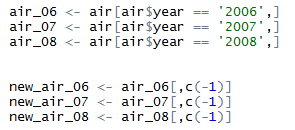




🡪주석참조

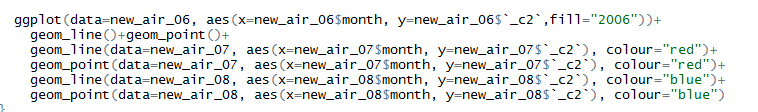


🡪년도별,월별,월별 딜레이 카운터를 출력하여 air객체에 담는다.

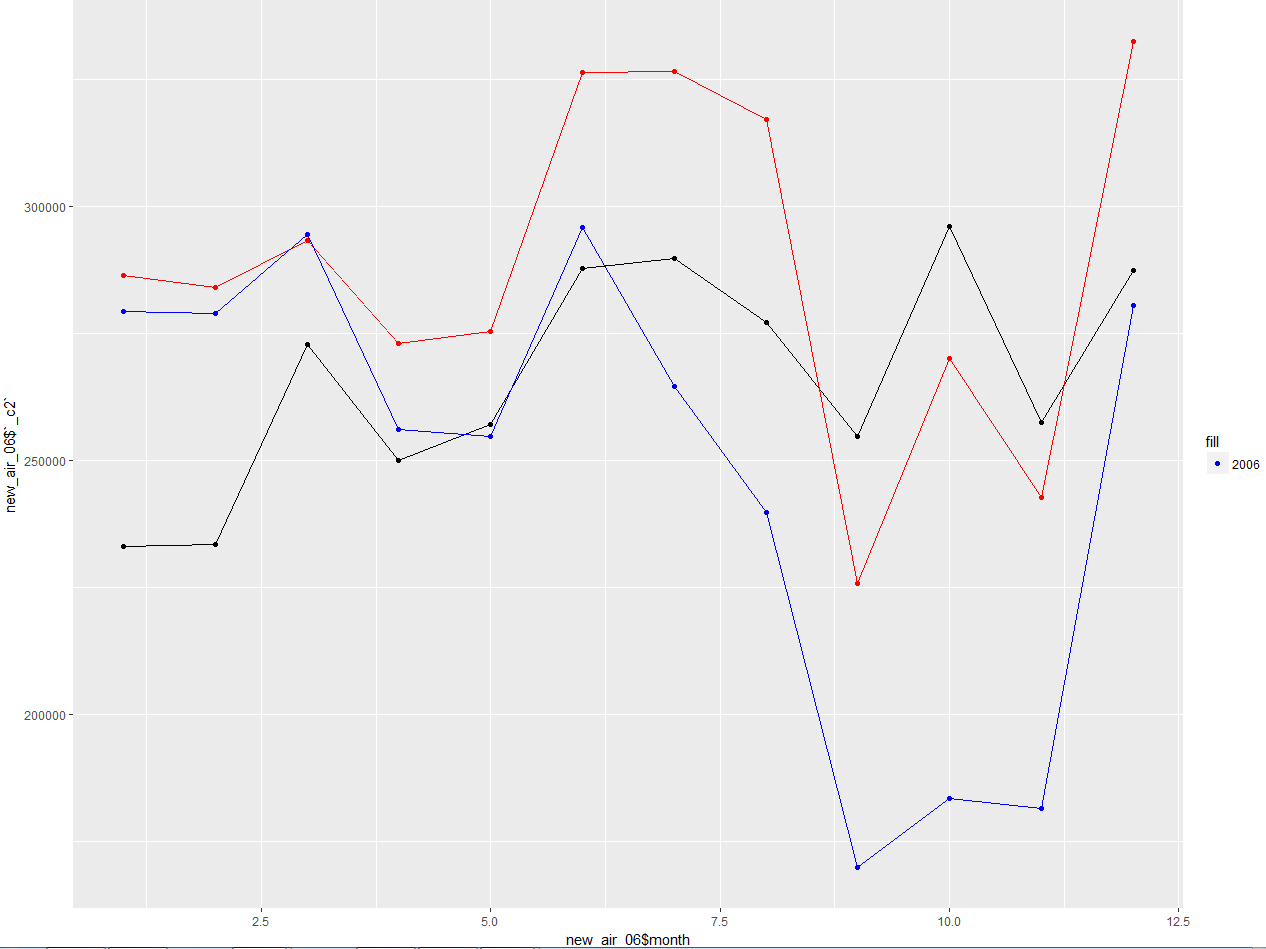


🡪연도,월,딜레이 count 칼럼 dataFrame을 만든 후, 각 연도에 따른 월, 딜레이 count칼럼을 dataFranme을 추출

4. 데이터 분석 화면을 구현 한다.



🡪그래프를 그려준다 geom\_line과 point를 사용하여 한층 보기쉽게 작성



🡪해석: 8월~9월달은 대체로 delay비율이 작고 1월~7월 까지 delay가 잦다. 이 그래프를 통해 알수 있는점은 겨울에 폭설로 인해 딜레이가 잦은것으로 보인다. 그에 반해 여름의 delay비율비 적은 것으로 추측된다.