# Quiz

1. B - The same PCollection can be written to multiple data sinks including BigQuery and Bigtable
2. A - Send the bad data into the DoFn as a SideOutput
3. A - Use a simple REST endpoint to trigger the pipeline
4. C - Create a composite key to group by multiple properties with GroupByKey
5. B - Use a SideInput to a ParDo
   a. CoGroupBy is shuffle heavy, a lot of network traffic between nodes and IO pressure
   b. Should not do a full outer join with ParDo, because the result will be too big

# Demo

Follow the guide here:
https://github.com/cs327e-fall2019/snippets/wiki/Jupyter-Notebook-with-Beam-and-Dataflow
If you get a 'permission denied' error add the --user flag to the command

```
pip2 install --user --upgrade virtualenv
```

Clone the snippets repo to the root directory on the notebook terminal
Open up the notebooks found in the repo
Replace the credentials with your project credentials and run the code