**Vision Document for Green Space VR Analytics**

# Introduction

## Purpose

The purpose of this vision document is to describe the underlying need for developing the Green Space VR (GSVR) Analytics software, the goals and needs of stakeholders, and the software products and features that will be developed to meet those needs. The vision document will be shared with stakeholders to help identify discrepancies between stakeholder and developer visions of an ideal green space virtual reality software product, and serve as the basis for developing use-cases and software requirements.

## Scope

This vision document is part of the GSVR Analytics software project. This project was created by Andrew Larkin to advance analytical methodologies for greenspace research and partially fulfill the Johns Hopkins University Introduction to GPU Programming (EN.605.417.81.FA17) and Social Media Analytics (EN.605.433.81.FA17) course requirements.

## Definitions, Abbreviations, Acronyms, and Abstractions

* GSVR: Green Space Virtual Reality
* VR: Virtual Reality
* MU: Multiple User
* OS: Operating System
* GPGPU: General Purpose Graphics Processing Unit
* OpenCL: Open Computing Language
* OpenGL: Open Graphics Library
* AMD: Advanced Micro Devices
* Req: Requirements

# Positioning

## Problem Statement

Natural components of urban environments (green space) are associated with multiple positive health outcomes, including increased attention restoration and cognition, decreased stress and depression, and increased sense of community. Scientists have hypothesized multiple pathways through which greenspace may influence health, but interactions between social, environmental and biological mechanisms of action make it difficult to estimate the relative importance of each pathway and identify specific pathways to promote for urban planning and public health. To further complicate matters, greenspace and human interactions with greenspace are dynamic and dependent on the time of day, week, and season, societal influences, and geographical region.

Novel data analytics and visualizations are needed to advance green space epidemiology and green space-mediated public health interventions. The purpose of this project is to develop an automated framework for collecting, analyzing, and visualizing green space related data from diverse data streams in real time using massively parallel general purpose graphics processing unit (GPGPU) analytics combined with virtual reality interfaces for enhanced data visualization.

## Product Position Statement

Several virtual reality data visualization project are in development. Promising projects include Alyssum and DatavizVR. While embracing the virtual reality hardware, these projects are not realizing the true potential of a three dimensional analytics platform. The Alyssum project to date is the only VR project that has shown a working prototype to the public. Alyssum uses three dimensions to provide expanded room for multiple 2-d graphs, resulting in software that would better be described as pseudo three dimensional rather than truly 3d. The Alyssum project is designed for single users, missing the potential that a simultaneous multiple user (MU) VR platform can provide for geographically dispersed collaborators. Both Alyssum and DatavizVR have limited utility for public health research as they do not attempt to provide novel means for integrating diverse data streams such as social media, satellite imagery, and smartphone measures, into a platform that facilitates comprehensive data integration, analysis and inference. Finally, both projects are not open source or free for non-profit research. The proposed GSVR presents a unique opportunity to provide a free open source tool that meets unfulfilled needs of the public health non-profit research and public health intervention communities. While this prototype project will focus on green space-related research, the developed code and methodologies can be utilized in many public health applications.

# Stakeholder and User Descriptions

## Stakeholder Summary

Stakeholders consist of all non-profit researchers and organizations who intend to support virtual reality informatics for non-profit research. Specific types of stakeholders include:

* Program users
  + Non-profit public health researchers
  + Public health professionals
  + Citizen Scientists
  + Policy makers
* Non-Profit Institutions
  + Academic
  + Public Health Agencies
* Software Developers
* Data stream providers
  + Social Media Outlets
  + Internet providers
  + Government Agencies
  + Google

## User Summary

## Primary users will consist of individuals who intend to utilize the GSVR program for research and public health interventions. These users will mostly consist of public health researchers, public health professionals, and policy makers who make decisions about public health intervention programs. As VR becomes more commonplace in the general public, citizen scientists will contribute to the primary user group.

## 

## User Environment

Unlike other data visualization platforms, when using virtual reality users are completely removed from visual inputs of their local environment. We expect the software to be utilized in environments where the user is stationary, either sitting or standing, and without obstructive movement from individuals or objects who are not actively participating in the virtual environment. Indoor noise levels may significantly vary. The ideal user environment will consist of an office space and desk. Most users will run the program for an hour or less at a time, with occasional multi-hour sessions.

Computer hardware environments will consist of computers that meet virtual reality requirements as described by NDVIDIA and AMD. Requirements include a core i5 or better processor, video card capable of running VR programs (see NVIDIA and AMD websites for VR-approved graphics cards) and 8gb of RAM. Users will also need to have either an Oculus Rift Controller, HTC Vive Controller, or Xbox Controller (available online for $30).

Computer software environments will consist of the Windows 10 OS or greater, and the GSVR executable program. Users will also need to have an active internet connection and administrative privileges for multiple user VR sessions. Programs that are likely to be running concurrently with GSVR include SteamVR, Oculus Rift Store, internet browsers Mozilla and Google Chrome, cloud storage programs Dropbox and Google Drive, Notification apps for Twitter, Facebook, Google Mail, and Microsoft Outlook, and security software including Norton Antivirus and Windows Defender.

## Stakeholder (non-user) profiles

Non-user stakeholders include:

* [NP] Nonprofit Institutions: NPs will provide hardware, approve and install the GSVR program for the majority of primary users. NPs may also determine whether GSVR can be granted internet access for multiple user sessions.
* [SD] Software Developers: SDs will modify, maintain, update, and improve both local and source versions of the GSVR code.
* [DP] Data Stream Providers: DPs consist of the numerous public and business organizations that provide publically available data to be integrated into the backend GSVR data analytics. DPs will be diverse and include multiple types of data streams, including social media, georeferenced satellite and non-satellite imagery, news reports, census data, environmental monitoring, and crowdsourcing.

## User profiles

Principle user stakeholders are defined as:

* [HR] Health Researchers: HRs will utilize GSVR to describe, analyze, and infer relationships between greenspace, health outcomes, and pathways of mediation. These users will be familiar with statistical analysis and greenspace research. HRs will have limited prior knowledge and experience with VR and computer programming. HRs will have specific research questions and populations of interest.
* [PP] Public Health Professionals: PPs will utilize GSVR to evaluate relationships between greenspace, health outcomes, and potential intervention strategies of interest. Familiarity with statistical analysis and greenspace research will vary more among PPs compared to HRs. HRs will also have limited experience with VR and computer programming, and will also have specific interventions and populations of interest.
* [CS] Citizen Scientists: CSs will be early adopters of utilizing VR for public health outreach and awareness. Compared to HRs and PPs, CSs will be more likely to have prior VR experience but less experience of statistical analyses and greenspace research. CSs will most likely be interested in utilizing GSVR for infographics and interventions.
* [PM] Policy Makers: PMs will utilize GSVR to evaluate potential interventions proposed by PPs. Potential uses of GSVR by PMs will include evaluating 3d infographics and making decisions based on PP recommendations. PMs will have significant experience making intervention-related decisions based on PP reports.

## Stakeholder Goals/Needs

|  |  |  |  |
| --- | --- | --- | --- |
| **Stakeholder** | **Needs/Goals** | **Priority** | **Proposed Solution** |
| All | Runs with minimal risk to hardware, other programs, and OS | 1 | Runs as an executable with no external packages and no writes outside of the internal SQL databases. |
| Users | Allow multiple users to participate and interact in VR sessions, including audio | 2 | Include an option for up to 4 simultaneous VR participants. Provide support for a listen server. |
| HR,PP | Select which data sources are shown | 1 | Provide UI interaction options that dynamically alter which datasets are analyzed by the underlying GPGPU algorithms |
| Users | Provide intuitive controls for all supported interfaces | 2 | Follow Epic Games best practices for VR controller development. |
| Users | Users can change time range and/or geographical region of presented data | 2 | Use third dimension to visually depict higher dimensional data, and add gesture recognition and UI interactions for geographical and temporal data range selection |
| Users | Provide predictions for social media sentiment, discussed topics, actions taken by community and influential stakeholders | 1 | Create machine learning models based on techniques that will be covered in Social Media Analytics course |
| HR, PP | Provide visual depiction of machine learning model performance and predictions | 2 | Include visual interactive graphs of model performance |
| Users | Include remote sensing data streams | 2 | Use Blender and QGIS to produce static meshes for georeferenced data visualization. |
| Users | Include social media data streams | 2 | Collect, filter and analyze data from Twitter, Facebook, YouTube, and Flickr using previously developed scripts and methods covered in Social Media Analytics course |
| HP,PP | Include custom user georeferenced data streams | 3 | Use Blender and QGIS to produce static meshes for georeferenced data visualization. |
| Users | Include georeferenced Google Street View (GSV) Imagery | 2 | Import GSV images for representative locations and add to virtual realy landscape. |
| Users | Include GSV imagery after applying a green screen algorithm | 3 | Apply the previously developed green screen algorithm to create screened GSV images. |
| Users | Users can find help and documentation | 1 | Provide in program instructions and guides. Write a user guide and include in the software package and .github site |
| SD | Code is easy to maintain and update | 2 | Follow Epic Games best practices. Provide documentation and source control via GitHub |

# Product Overview

## Product Perspective

GSVR will run on Windows System and be dependent on a VR headset for output and external controller for input. SQL databases will be used for data storage. Data processing algorithms will be written in c++, with GPGPU acceleration using OpenCL. The Virtual reality interface will be written in c++ using the Unreal Engine. The program will require an internet connection for multiple user VR sessions and real time data analytics. The program will be functional for single user VR and historical data analytics without an internet connection. The program will retain user preferences and will cache frequently accessed summary statistics.

## Summary of Capabilities

|  |  |
| --- | --- |
| **Customer Benefit** | **Supporting Features** |
| Users can comprehensively evaluate relationships between greenspace, mediating factors, and health outcomes | Real time data visualization, GPGPU-based machine learning and data analytics, expert-based UI design for diverse data stream integration, real time model performance scores and model predictions |
| Analytics are easier for geographically dispersed collaborations | Multiple player real time interactive VR sessions, audio communications. |
| The platform can be expanded to include additional data streams or exposures of interest | Open source code, modular design, .github directory for community improvements, published software engineering documents and software development blog. |

## Assumptions and Dependencies

* Users are capable of operating a Windows environment and have or can obtain authorization to run an executable file
* Users can read English at a high school level or above
* Users are familiar with the general concept of virtual reality
* Users have knowledge equivalent to an undergraduate course in statistics
* Users will be able to manipulate their external controller of choice
* Users are not overly sensitive to virtual reality sickness
* Computer Programmers have experience developing for the Unity or Unreal Engine
* Epic Games will continue to provide Unreal Engine source code for free for non-profit applications
* Users who wish to use the multi user functionalities will have broadband or DSL access to the internet

## Cost and Pricing

GSVR will be free for all users.

## Licensing and Installation

GSVR will be provided under the MIT License, which provides the program as open source and available for free, contingent that any products that utilize the GSVR code will provide acknowledgement of the code source and also provide derived products as free and open source.

# Constrains

* Milestones and deliverables are based on course deadlines
* Development hours are limited to weekends and finals week
* No budget - resources are limited to open source programming environments and free data streams
* Data streams will further be limited to publicly available data
* Geographical region for the prototype will be limited to the Continental United States